Building Development Monitoring in Multitemporal Remotely Sensed Image Pairs with Stochastic Birth-Death Dynamics

Csaba Benedek, Xavier Descombes, Josiane Zerubia

Building Development Monitoring in Multitemporal Remotely Sensed Image Pairs with Stochastic Birth-Death Dynamics

Csaba Benedek, Xavier Descombes and Josiane Zerubia *Fellow, IEEE*

Abstract—In this paper we introduce a new probabilistic method which integrates building extraction with change detection in remotely sensed image pairs. A global optimization process attempts to find the optimal configuration of buildings, considering the observed data, prior knowledge, and interactions between the neighboring building parts. We present methodological contributions in three key issues: (1) We implement a novel object-change modeling approach based on Multitemporal Marked Point Processes, which simultaneously exploits low level change information between the time layers and object level building description to recognize and separate changed and unaltered buildings. (2) To answering the challenges of *data heterogeneity* in aerial and satellite image repositories, we construct a flexible hierarchical framework which can create various building appearance models from different elementary feature based modules. (3) To simultaneously ensure the convergence, optimality and computation complexity constraints raised by the increased *data quantity*, we adopt the quick *Multiple Birth and Death* optimization technique for change detection purposes, and propose a novel non-uniform stochastic object birth process, which generates relevant objects with higher probability based on low-level image features.

Index Terms—Building extraction, change detection, Marked Point Processes, Multiple Birth and Death Dynamics

1 Introduction

Following the evolution of built-up regions is a key issue of aerial and satellite image analysis. Although the topic has been extensively studied since the 80's, it has had to continuously face the challenges of the quickly evolving quality and quantity of remotely sensed data, the richness of different building appearances, the data-heterogeneity in the available image repositories and the various requirements of new application areas.

1.1 Input Data

Numerous methods in the bibliography address building extraction at a single time instance [1], [2], [3]. It is common to use multiview inputs [4], [5] to exploit 3-D information in building modeling. Detection in densely populated areas can be efficient by working on stereo- or lidar-based Digital Elevation/Surface Models (DEM/DSM), where the silhouettes of the building footprints can be separated from the ground planes by the estimated height data [2], [6], [7], [8]. Other benefits are provided by multiple sensor inputs such as fusion of aerial images with color infrared (CIR) [9], or laser data [10]. However, several image repositories from city suburbs and smaller settlements lack stereo or special sensor information. We address this case in the paper: building identification becomes here a challenging monocular object recognition task based on purely optical data [11].

- *C. Benedek is with the Distributed Events Analysis Research Group, Computer and Automation Research Institute, H-1111 Kende utca 13–17, Budapest, Hungary. E-mail: bcsaba@sztaki.hu His present work was partially funded by the INRIA postdoctoral fellowship, in the Ariana Project Team, INRIA Sophia Antipolis–Méditerrannée, France, and by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.*
- *X. Descombes and J. Zerubia are with the Ariana Project Team, INRIA Sophia Antipolis–Méditerranée, 2004 route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex, France. E-mail: {xavier.descombes, josiane.zerubia}@inria.fr*

1.2 Multitemporal Information

Remote sensing image databases often contain multitemporal image samples from the same geographical areas. Exploiting the temporal information, change recognition and classification are nowadays important aspects of urban scene analysis. Several recent building change detection approaches [7], [12] assume that a topographic building database is already available for the earlier time layer, thus the process can be decomposed into old model verification and new building exploration phases. On the other hand, when dealing with image repositories without any meta data, the task requires automatic building detection for each image.

In this paper, we solely use as input a registered pair of (2-D) images taken at several years time difference. Applying conventional stereo-matching algorithms in this case may face several difficulties. First, the scene content, the viewpoints and the image qualities of the two views may be significantly different, which can corrupt feature matching algorithms (e.g. corner point tracking) needed for 3-D structure extraction. Moreover, several databases contain images which are created by mosaicking separately taken aerial photos, and the components undergo different geometric corrections.

In the proposed approach we apply a 2-D building and change detection technique [13] which is less influenced by the above effects than stereo based approaches. Moreover, at a higher processing level, the proposed footprint extraction step may also contribute to 3-D building reconstruction [2], [6].

Change detection methods frequently rely on the assumption that changes occur very rarely, thus they can be identified through outlier detection using global image statistics [14]. However, in dynamically improving (sub-)urban areas this hypothesis is often invalid, and there is a need for solutions which are insensitive to the quantity of differences.

An object oriented change detection technique is introduced in [13] and applied to the extraction of damaged buildings after natural disasters. This method follows the Post Detection Comparison (PDC) approach, as independent building detection processes are applied for the two images, followed by object level comparison. However, the object detection phase can be corrupted by image noise, irregular structures or occlusions by vegetation [13] which may present missing or only partially extracted buildings to the object matching module. Moreover, this comparison may be affected by further intensity artifacts caused by shadows or altered illumination conditions.

Following another approach, several low level change detection methods have been proposed for remote sensing [15], [16], which work without using any explicit object models. They extract image regions which have been altered in an irregular way based on an appropriately selected set of features, such as color difference, texture or block correlation. Although these techniques are usually considered as preprocessing filters, there have not been many attempts given to justify how they can support the object level investigations. We take a step forward in this paper, and exploit interaction between object extraction and local textural image-similarity information in a unified probabilistic model. It will be shown that we can obtain additional evidences for the presence of new, modified or demolished buildings through detecting changes in relevant low level feature domains. As for *unchanged* buildings, the images of the two time instances provide multiple views of the same objects, which may increase the detection accuracy compared to relying on a single time layer.

1.3 Object and Configuration Models

Another important issue is related to modeling the building entities. The conventional *bottom-up* techniques [17] construct the objects from primitives, like roof blobs, edge parts or corners. Although these methods can be fast, they may fail if the primitives cannot be reliably detected. To increase robustness, it is common to follow the Hypothesis Generation-Acceptance (HGA) scheme [3], [18]. Here the accuracy of object proposition is not crucial, as false candidates can be eliminated in the verification step. However, objects missed by the generation process cannot be recovered later, which may result in several false negatives. On the other hand, generating too many object hypotheses (e.g. applying exhaustive search) slows down the detection process significantly. Finally, conventional HGA techniques search for separate objects instead of global object configurations, disregarding population-level features such as overlapping, relative alignment, color similarity or spatial distance of the neighboring objects [2].

To overcome the above drawbacks, recent *inverse methods* [19] assign a fitness value to each possible object configuration, and an optimization process attempts to find the configuration with the highest confidence. This way, flexible object appearance models can be adopted, and it is also straightforward to incorporate prior shape information and object interactions. Marked Point Processes (MPP) [19] are good candidates for addressing these challenges, since they can efficiently model the geometry of objects and deal with an unknown number of entities [6], [20], [21]. However, this inverse approach needs to perform a computationally expensive search in a high dimensional population space, where local maxima of the fitness function can mislead the optimization. Due to the large databases, the optimization issue plays a particular role in remote sensing applications. In previous techniques [6], [20], [21] the optimization has been performed using a Reversible Jump Markov Chain Monte Carlo (RJMCMC) scheme, with implementations where each iteration perturbs one or a couple of objects, and the rejection rate, especially for the birth move, induces a heavy computation time. Besides, one should be very careful when decreasing the temperature, because at low temperature, it is difficult to add objects to the population.

Taking a different approach, we adopt here the Multiple Birth and Death Dynamic technique (MBD) [22] for the change detection purposes. Unlike following a discrete jump-diffusion scheme like in RJMCMC, the MBD optimization method defines a continuous time stochastic evolution of the object population, which aims to converge to the optimal configuration. The evolution under consideration is a birth-and-death equilibrium dynamics on the configuration space, embedded into a Simulated Annealing (SA) process, where the temperature of the system tends to zero in time. The final step is the discretization of this non-stationary dynamics: the resulting discrete process is a non-homogeneous Markov chain with transition probabilities depending on the temperature, energy function and discretization step. In practice, the MBD algorithm evolves the population of buildings by alternating purely stochastic object generation (*birth*) and removal (*death*) steps in a SA framework. In contrast to the above RJMCMC implementations, each birth step of MBD consists of adding *several* random objects to the current configuration, which is allowed due to the discretization trick. Using MBD, there is no rejection during the birth step, therefore high energetic objects can still be added independently of the temperature parameter. Thus the final result is much less sensitive to the tuning of the SA temperature decreasing process, which can be achieved faster. Due to these properties, in selected remote sensing tasks (bird and tree detection) [22] the optimization with MBD proved to be around ten times faster than RJMCMC with similar quality results. In addition, MBD has already been applied in different application areas, such as cell counting [23] and video surveillance [24].

Another key point is the probabilistic approach for object proposal. In several previous MPP applications [6], the generation of object candidates followed prior (e.g. Poisson) distributions. On the contrary, we apply a data driven birth process to accelerate the convergence of MBD, which proposes relevant objects with higher probability based on various image features. In addition, we calculate not only a probability map for the object centers, but also estimate the expected object appearances through low-level descriptors.

This approach uses a similar idea to the Data Driven MCMC scheme of image segmentation [25]. However, while in [25] the *importance proposal probabilities* of the moves are used by a jump-diffusion process, we should embed the data driven exploration steps into the MBD framework.

In this paper, we propose a novel *multitemporal MPP* (mMPP) model and an efficient *bi-layer MBD* (bMBD) optimization algorithm for the 2-D building change detection problem in remotely sensed image pairs. The present approach has been partially introduced in [26], [27], and further demonstrating figures and experimental results are provided in [28]. Due to its modularity, the method could be easily adapted to different object level change detection applications, for instance tree or road detection. On the other hand, we attempt to focus on the task specific issues as well. We present a broad feature library, which can be appropriate for the detection of a large set of buildings, expecting various image properties. For this reason, in the following section we provide an overview on the state-of-the art methods for monocular building extraction.

1.4 Related Works in Monocular Building Detection

A *SIFT* key point based method has been presented in [29] for urban area extraction and building detection. This technique assumes that the building structures in a given image can be efficiently characterized by a couple of template buildings (here two templates: a bright and a dark one) which are used for training. However, images containing a high variety of buildings may need a huge template library, where the overlap between the building and background domains in the descriptor space may be hard to control. A recent model based on Gabor filters (*Gabor*) [30] represents building positions in the image as joint probability density functions of four different local feature vectors, and performs data and decision fusion in a probabilistic framework to detect building locations. It is important to note that in [29] and [30] the goal is building localization, but the roof outlines are not extracted, which makes it difficult to apply the method for change detection.

A stochastic *MRF* framework is introduced in [1] for detecting building rooftops from single images, which combines 2-D and 3-D information. This approach is based on hierarchical grouping of extracted edge segments to form continuous lines, junctions and finally closed curve hypotheses. However, several restrictions are applied for buildings: it is assumed that they have uniform height, they are composed of planar surfaces with parallel sides, and each building casts its shadow on a locally flat surface. Similarly to [18], [31] the method needs a reasonable edge map, because missing large side parts and false edges inside and around the buildings may corrupt the edge grouping process. Edge based building detection is also applied in [32] as a part of a complete scene interpretation process. This approach deals with different object categories, and implements multi-level interactions within a scene and between different object types as well.

Combining roof color, shadow and edge information has been suggested in [3] in a two-step process which we refer to later as the Edge Verification (*EV*) approach. In EV, color and shadow are used first for coarse built-in candidate area estimation; thereafter, fitting the building rectangles and verification of the proposals are based purely on the Canny edge map of the obtained candidate regions. As a drawback, this sequential approach is sensitive to the failure of each individual feature. A corrupted edge image causes unreliable corner detection and edge mask stretching, meanwhile, without shadow and color information, the building search area should be extended for the whole image, increasing the processing time and the appearance of false edge patterns.

Segment-Merge (*SM*) techniques follow an approach different from edge based methods, as they consider building detection as a region level or image segmentation problem [17], [33], [34]. In [34] the authors assume that buildings are homogenous areas w.r.t. either color or texture, which can be used for training-based background subtraction. Hence, elementary constraints for shape and size are used to group the candidate regions into building objects. This method can fail, if the background and building areas cannot be efficiently separated with the chosen color or texture descriptors, thus several building and background parts are merged in the same regions of the oversegmented map. On the other hand, for homogenous buildings (see BEIJING, Fig. 4) or salient roof colors (see BUDAPEST red roofs, Fig. 26, top) region features are often more robust than weak or ragged edge maps.

Beside probabilistic models [2], [6], variational techniques [35], [36] have been proposed recently for building extraction through energy minimization. Similarly to our method, the Recognition-Driven Variational (RDV) framework of [35] is based on *data* and *prior* term decomposition. However, they focus principally on the prior shape modeling issue and use a simplified image-dependent model part, which assumes that the building and background regions can be roughly separated through considering them as locally homogenous intensity classes. In cases where this data term cannot detect probable building regions, the algorithm naturally fails.

From another point of view, the *prior models* of [2], [35] contain libraries of complete object shapes, while other approaches [1], [6] construct the objects from elementary building blocks (rectangles or line segments), and the higher level shape information is encoded by interaction constraints of the nearby components. While global description of RDV [35] can be efficient if all objects of the scene can be characterized by a restricted number of prototype shapes, the algorithm fails to detect the boundaries accurately, if a given building cannot be sufficiently represented by any shape from the database, using any possible planar projection. On the other hand, the constructive approach - which we follow in the current paper - is preferable if the prior models of the buildings are partially unknown or largely diverse.

As for *image data modeling*, the above overviewed methods are based on image- or scene-specific hypotheses, such as unique roof colors [29], shadows [1], [3], strong edges [1], [3], [18], [31], [32], homogeneous roofs [17], [33], [34], or a limited number of 2-D [29] or 3-D [2] building templates. The obvious limitations of these techniques come from the nature of the varying image data, and the lack of adaptivity to different circumstances. To develop more generic models, besides the extraction of the descriptors, feature integration

Fig. 1. Definition of the rectangle parameters

and selection should be addressed at the same time. Therefore we construct a framework which can combine the features in a flexible way depending on availability, accommodating an extended set of images and situations.

2 Problem Definition

The input of the proposed method consists of two co-registered aerial or satellite images which were taken from the same area with several months or years of time difference. Thus a single photo is available at each time instance, and we cannot exploit additional meta-information such as maps or topographic building databases. We expect the presence of registration or parallax errors, but we assume that they only cause distortions of a few pixels. We consider each building to be constructed from one or many rectangular building segments, which we aim to extract by the model described in the following. As output we provide the size, position and orientation parameters of the detected building segments, and give information which objects are new, demolished, modified/rebuilt or unchanged.

Let us denote by S the common $S_W \times S_H$ pixel lattice of the input images and by $s \in S$ a single pixel. Let u be a building segment candidate assigned to the input image pair, which is jointly characterized by geometric and temporal attributes. We consider the center of each building, $c = [c_x, c_y]$ as a point in $[0, S_W] \times [0, S_H] \subset \mathbb{R}^2$, which can be projected to S by simple discretization: $c \rightarrow [\lfloor c_x \rfloor, \lfloor c_y \rfloor]$. Let the rectangle $R_u \subset S$ be the set of pixels corresponding to u. Apart from the center, R_u is described by the e_L, e_l side lengths, and $\theta \in [-90°, +90°]$ orientation parameters as shown in Fig. 1.

For purposes of dealing with multiple time layers, we assign to each u an index flag, $\xi(u) \in \{1, 2, *\}$, where '$*$' indicates an unchanged object (i.e. present in both images), while '1' and '2' correspond to building segments which appear *only* in the first *or* second image respectively. We will denote the set of all the possible object records u=(c_x, c_y, e_L, e_l, θ, ξ) by $\mathcal{H}$. The output of the proposed model is a configuration of building segments, $\omega \in \mathcal{H}^n$, where n, the number of objects is also unknown.

The method exploits rough preliminary knowledge about the object sizes, which will be introduced in two steps for easier interpretation. In the first part of the discussion, we assume that the side length parameters of the building segments in the scene have the same order of magnitude and can be constrained by $e_L(u) \in [e_L^{\min}, e_L^{\max}]$ and $e_l(u) \in [e_l^{\min}, e_l^{\max}]$. Later in Section 5 we present a multi-scale extension of the process, which enables us to handle of image inputs which contain buildings with significantly different sizes.

3 Feature Selection

In this section, we introduce different image features for building and change recognition. Since the proposed model obtains the optimal object configuration through stochastic birth-death iterations, two essential questions should be answered based on the image data. First, how can we efficiently generate relevant objects during the *birth* process? Secondly, how can it be ensured that the adequate objects survive the *death* step? To keep focus on both challenges, we utilize low level and object level features in parallel.

Low level features are extracted around each pixel as typical color, texture and local similarity between the time layers. They are principally used in the birth step, to estimate where the buildings *might* be located, what they *might* look like, and where changes *should* be expected. As a consequence, objects are generated with higher probability in the estimated built-up regions, considering the estimated appearance models.

On the other hand, object level features evaluate a building hypothesis for each proposed oriented rectangle. The choice of preserving *or* killing an object in the *death* step strongly depends on object descriptors, thus their accuracy is crucial.

3.1 Low level features for building detection

We begin the discussion with low level features extracted from individual images. For the purposes of built-in area estimation, at each pixel s we calculate a pair of birth probabilities, $P_b^{(1)}(s)$ and $P_b^{(2)}(s)$, which give the likelihood of s being an object center in image 1, and 2, respectively. The nomination refers to the fact that in the *birth* step the frequency of proposing an object at s will be proportional to the local birth probabilities. On the other hand, we also assign expected orientation $\mu_\theta^{(i)}(s)$, and side length values $\mu_L^{(i)}(s)$ resp. $\mu_l^{(i)}(s)$ to the image pixels, which help in estimating the θ, e_L and e_l parameters of objects centered at s based on various descriptors from the i^{th} image ($i \in \{1, 2\}$). Since the calculation of birth, orientation and mean side length maps are the same for both time layers, we simplify the notation by ignoring the image index in the following part of this section. Later on, we will denote the time stamp again by a superscript index in parentheses wherever necessary.

3.1.1 Local Gradient Orientation Density

The first feature exploits the fact that regions of buildings should contain edges in *perpendicular* directions. This property can be robustly characterized by local Gradient Orientation Density Functions (GODF) [37]. Let ∇g_s be the intensity gradient vector at pixel s with magnitude $||\nabla g_s||$ and angle ϑ_s^∇. Let $W_l(s)$ be the rectangular $l \times l$ sized window around s, where l is chosen as $W_l(s)$ can cover an average building from the training set narrowly. For each s we calculate the weighted ϑ_s^∇ density of $W_l(s)$:

$$\lambda_s(\vartheta) = \frac{1}{N_s} \sum_{r \in W_l(s)} \frac{1}{h} \cdot ||\nabla g_r|| \cdot k\left(\frac{\vartheta - \vartheta_r^\nabla}{h}\right) \qquad (1)$$

where $N_s = \sum_{r \in W_l(s)} ||\nabla g_r||$, and $k(.)$ is a kernel function with a bandwidth parameter h. We use uniform kernels for

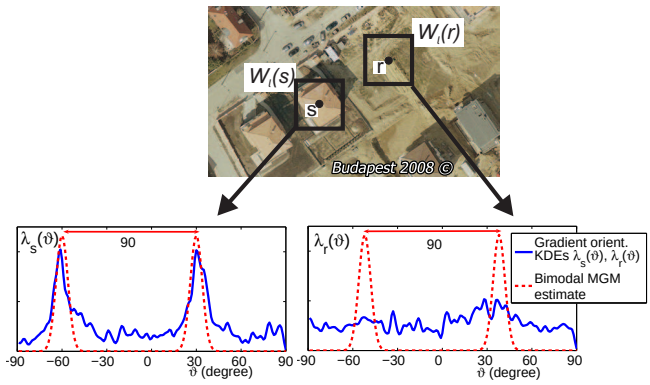
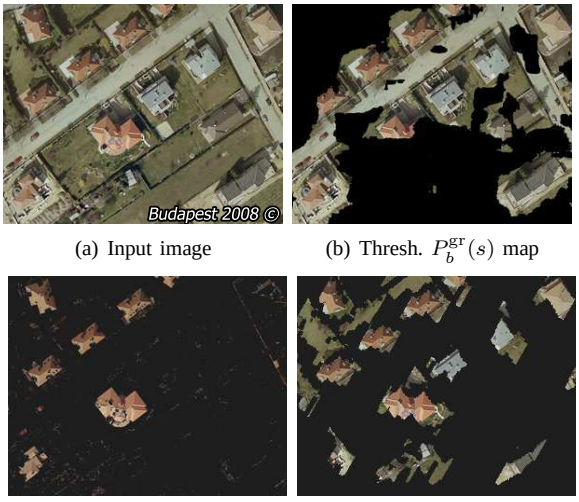


Fig. 2. Kernel density estimation of the local gradient orientation histogram around two selected pixels: a building center s and an empty site r.



(a) Input image (b) Thresh. $P_b^{\mathrm{gr}}(s)$ map

(c) Color mask ($\varrho_{\mathrm{co}}(s) = 1$) (d) Thresh. $P_b^{\mathrm{sh}}(s)$ map

Fig. 3. Building candidate regions obtained by the low level (b) gradient (c) color and (d) shadow descriptors

quick calculation. If $W_l(s)$ covers a building, the $\lambda_s(\vartheta)$ function has two peaks, located at a distance of 90° from each other in the ϑ-domain (see Fig. 2). This property can be measured by correlating $\lambda_s(\vartheta)$ with an appropriately matched bi-modal density function:

$$\alpha(s,m) = \int \lambda_s(\vartheta)\eta_2\left(\vartheta, m, d_\lambda\right) d\vartheta \qquad (2)$$

where $\eta_2(.)$ is a mixture of two Gaussians with mean values m, resp. $m+90^\circ$, and deviation d_λ for both components (d_λ is a parameter of the process set by training). Offset m_s and value α_s of the maximal correlation can be obtained as:

$$m_s = \underset{m\in[-90^\circ,0]}{\operatorname{argmax}} \left\{\alpha(s,m)\right\} \qquad \alpha_s = \alpha\big(s, m_s\big)$$

Pixels with high α_s are more likely to be centers of buildings, which can be encoded in a gradient-based birth map $P_b^{\mathrm{gr}}(s) = \alpha_s / \sum_{r\in S}\alpha_r$. For the sample image in Fig. 3, a thresholded P_b^{gr} map is shown in Fig. 3(b).

Furthermore, let us observe that offsets m_s and $m_s + 90^\circ$ estimate the dominant gradient directions in the $W_l(s)$ region. Thus, for a building with center s, we expect its θ parameter around a mean orientation value $\mu_\theta(s)$, defined as:

$$\mu_\theta(s) = \begin{cases} m_s & \text{if } \lambda_s(m_s) > \lambda_s(m_s+90^\circ) \\ m_s + 90^\circ & \text{otherwise} \end{cases} \qquad (3)$$

For this reason if the birth step proposes an object u at pixel s, its orientation is set as $\theta(u) = \mu_\theta(s) + \eta_\theta$, where η_θ is a random value, generated for each object independently according to a zero-mean Gaussian distribution with a small deviation parameter σ_θ.

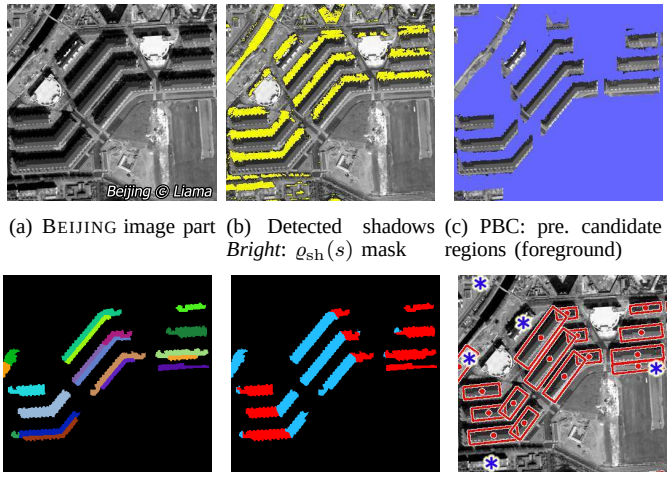
3.1.2 Roof color filtering and shadow evidence

Several types of roofs can be identified by their typical colors [3]. Let us assume that based on a roof color hypothesis, we extract an indicator mask $\varrho_{\mathrm{co}}(s) \in \{0,1\}$ (e.g. by thresholding a chrominance channel), where $\varrho_{\mathrm{co}}(s) = 1$ marks that s has roof color. Many roof pixels are expected around building centers, thus for each s we calculate the accumulated $\varrho_{\mathrm{co}}-$filling factor in its neighborhood: $\Gamma_s = \sum_{r\in W_l(s)} \varrho_{\mathrm{co}}(r)$. The color birth map value is obtained as $P_b^{\mathrm{co}}(s) = \Gamma_s / \sum_{r\in S}\Gamma_r$. Note that due to color overlapping between the roofs and the background [3], the $\varrho_{\mathrm{co}}(s)$ mask often only contains a part of the building segments (e.g. only *red* roofs are detected in Fig. 3(c)). Particularly, in grayscale images, the overlap between intensity domains of the classes is usually too large for any reasonable separation.

A supplementary evidence for the presence of buildings can be obtained through their *cast shadows* [1], [3]. In several types of remote sensing scenes, a binary shadow mask $\varrho_{\mathrm{sh}}(s)$ can be derived by filtering pixels from the dark-blue color domain [38]. The relative alignment of shadows to the buildings is determined by the global Sun direction, which can be set with minor user interaction or calculated automatically [3]. Consequently, we can identify the building candidate areas as image regions lying next to the shadow blobs opposing the *Sun direction* (see Fig. 3(d) and later Fig. 10). As for the shadow based birth map, we use a constant birth rate $P_b^{\mathrm{sh}}(s) = p_0^{\mathrm{sh}}$ within the obtained candidate regions and a significantly smaller constant on the outside. It is also important to note that for building detection only the cast shadows (i.e. shadows on the ground) are relevant, while self shadows (i.e. weakly or not illuminated building parts) should be ignored. However, as pointed out in [39], in most cases cast and self shadows have different intensity values, since the shadowed object parts are mostly illuminated by secondary light sources such as reflections from surrounding buildings.

3.1.3 Roof homogeneity

As illustrated in Fig. 3, the $P_b^{\mathrm{gr}}(s)$ and $P_b^{\mathrm{sh}}(s)$ birth maps usually give a quite coarse estimation of the built-up regions, which is hardly appropriate for building separation and size estimation. Although we may obtain notably accurate footprints through roof color filtering (Fig. 3(c)), it can only be used for a limited subset of the images and objects. On the other hand, in high resolution images provided by satellites such as Ikonos and Quickbird, a significant part of the roof tops can be identified as homogeneous blobs in the coarsely detected



(a) BEIJING image part (b) Detected shadows *Bright*: $\varrho_{\mathrm{sh}}(s)$ mask (c) PBC: pre. candidate regions (foreground)

(d) HBC: large homogenous regions of PBC (e) GHBC: grad. orient. based clusters of HBC (f) Building rectangles based on GHBC

Fig. 4. Preliminary building estimation based on roof homogeneity. Missing and false alarms – denoted by $(*)$ in image (f) – are eliminated later in the process.

building candidate regions. In this section we investigate how *roof homogeneity* can be exploited for building region detection and refinement.

The feature extraction algorithm consists of the following steps (illustration for the BEIJING image is shown in Fig. 4):

- `Candidate Region Filtering`: for a given input image (Fig. 4(a)) obtain the coarse preliminary building candidate (PBC) regions based on the gradient and/or shadow features, as explained in Sec. 3.1.1 and 3.1.2 (Fig. 4(b)-(c)).
- `Intensity based segmentation`: we (over)-segment the PBC regions of the input image into homogenous components, and ignore the blobs smaller than 20% of the expected mean building area. This step results in the homogenous building candidate (HBC) region map (Fig. 4(d)).
- `Orientation based clustering`: we re-cluster the HBC map based on the $\mu_\theta(s)$ dominant local gradient orientation values obtained in the regions of interest, and call the result GHBC image as shown in Fig. 4(e). Each uniform component of GHBC is considered in the following as a building segment candidate.
- `Candidate parameter estimation`: we estimate the center and the bounding box (Fig. 4(f)) parameters for each building segment candidate through morphological box fitting techniques.

Let us denote the candidate rectangles (Fig. 4(f)) obtained in the previous filtering process by $\mathcal{R}_i$, $i = 1 \dots t$, and let $c(\mathcal{R}_i)$ be the center of $\mathcal{R}_i$. Then, for each pixel, we determine the closest rectangle $\mathcal{R}_s^{\min} = \arg\min_i ||s - c(\mathcal{R}_i)||$ and calculate the homogeneity birth value as:

$$P_b^{\mathrm{ho}}(s) = k_{\mathcal{R}}\left(\frac{||s - c(\mathcal{R}_s^{\min})||}{h_{\mathcal{R}}}\right) \tag{4}$$

with a $k_{\mathcal{R}}(.)$ kernel function, and $h_{\mathcal{R}}$ bandwidth parameter.

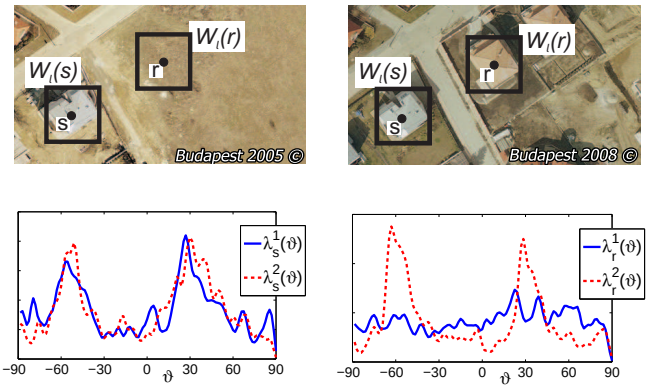


Fig. 5. Comparing the $\lambda(.)$ functions in the two image layers regarding two selected pixels. s corresponds to an unchanged point and r to a built-up change.

Besides marking the candidate regions of the building centers, the $\{\mathcal{R}_i | i = 1 \dots t\}$ set provides local estimations for the side length parameters: $\mu_L(s) = e_L(\mathcal{R}_s^{\min})$ and $\mu_l(s) = e_l(\mathcal{R}_s^{\min})$. Of course, we can only assume this information to be reliable in pixel positions with *high* homogeneity birth factors. Thus, for an object u proposed at s, we set the side length values with a probability proportional to $P_b^{\mathrm{ho}}(s)$ as:

$$e_L(u) = \mu_L(s) + \eta_L(s), \qquad e_l(u) = \mu_l(s) + \eta_l(s)$$

where $\eta_L(s)$ and $\eta_l(s)$ are independent zero mean Gaussian random variables. Note that side length estimates can be similarly extracted from the color feature map. This preliminary calculation is particularly significant if the object sizes show a large variety, since sampling the side length parameters of the proposed objects according to a prior distribution with a wide support can slow down the speed of the iterative birth-death process critically.

3.2 Low level change feature

Up to this point, we have used various descriptors to estimate the location and appearance of the buildings in the individual images. However, the gradient orientation statistics also offers a tool for low level region comparison, which can be directly involved in the scenario model. Let us consider the $\lambda_s(.)$ orientation density introduced in Sec. 3.1.1. Matching the $\lambda_s^{(1)}(.)$ and $\lambda_s^{(2)}(.)$ functions from the two time layers can be interpreted as low level similarity checking of the areas around s in the two images, based on "building-focused" textural features (see Fig 5), which are independent of illumination and coloring effects and robust regarding parallax and registration errors. For measuring the dissimilarities, we use the Bhattacharyya distance:

$$b(s) = -\log \int \sqrt{\lambda_s^{(1)}(\vartheta) \cdot \lambda_s^{(2)}(\vartheta)} d\vartheta \tag{5}$$

Choosing an appropriate b_0 threshold [14], the binarized pixel level change mask is obtained as:

$$\varrho_{\mathrm{ch}}(s) = \begin{cases} 1 & \text{if } b(s) > b_0 \\ 0 & \text{if } b(s) \leq b_0 \end{cases} \tag{6}$$

(a) Input image 1 (b) Input image 2 (c) Changed regions of image 2

Fig. 6. Low level change detection: (a) and (b) input images, (c) Bhattacharyya change mask $\varrho_{\rm ch}$

As shown in Fig. 6, the above comparison separates efficiently the image regions which contain the changed and unchanged buildings, respectively. Knowing that l^2 is the area of window $W_l(s)$, the probability of change around pixel s is derived as:

$$P_{\rm ch}(s) = \sum_{r \in W_l(s)} \varrho_{\rm ch}(r)/l^2 \tag{7}$$

Considering the change feature, we can exploit an additional information source for scene interpretation, which is independent of the object recognizer.

3.3 Integration of the different birth maps

Since the main goal of the *combined birth map* in each image is to keep focus on all building candidate areas, we derive it with the maximum operator from the birth maps of the features. For example, when gradient, color and shadow are simultaneously used, we obtain the final field as $P_b(s) = \max\left\{P_b^{\rm gr}(s), P_b^{\rm co}(s), P_b^{\rm sh}(s)\right\}\ \forall s \in S$. For input, without shadow or color information, we can ignore the corresponding feature in a straightforward way, or exchange the $P_b^{\rm co}(s)$ component to the homogeneity birth value, $P_b^{\rm ho}(s)$.

In the birth step of the bMBD process, the birth maps of both time layers, $P_b^{(1)}(s)$ and $P_b^{(2)}(s)$, and the change map $P_{\rm ch}(s)$ are utilized in parallel. We propose an unchanged object at s with a probability proportional to $(1 - P_{\rm ch}(s)) \cdot \max_{i\in\{1,2\}} P_b^{(i)}(s)$, while at the same location, the likelihood of generating a changed building segment is $P_{\rm ch}(s) \cdot P_b^{(i)}(s)$ for image i.

3.4 Object-Level Features

Besides efficient object generation, the second key point of the applied birth-death dynamics based approach is to validate the proposed building segment candidates. In this section, we construct a $\varphi^{(i)}(u) : \mathcal{H} \to [-1, 1]$ energy function, which calculates a negative building log-likelihood value of object u in the $i^{\rm th}$ image (hereafter we ignore the i superscript). By definition, a rectangle with $\varphi(u) < 0$ is called *attractive* object, and we aim to construct the $\varphi(u)$ function so that attractive objects correspond exclusively to the true buildings.

The process consists of three parts: feature extraction, energy calculation and feature integration. *First*, we define different $f(u) : \mathcal{H} \to \mathbb{R}$ features which evaluate a building hypothesis for u in the image, so that 'high' $f(u)$ values correspond to efficient building candidates. In the *second step*, we construct energy subterms for each feature f, by attempting to satisfy $\varphi_f(u) < 0$ for real objects and $\varphi_f(u) > 0$ for false candidates. For this purpose, we project the feature domain to $[-1, 1]$ with a monotonously decreasing function shown in Fig. 8: $\varphi_f(u) = \mathcal{Q}\big(f(u), d_0^f, D^f\big)$ where

$$\mathcal{Q}(x, d_0, D) = \begin{cases} \left(1 - \frac{x}{d_0}\right), & \text{if} \quad x < d_0 \\ \exp\left(-\frac{x-d_0}{D}\right) - 1, & \text{if} \quad x \geq d_0 \end{cases} \tag{8}$$

Observe that the $\mathcal{Q}$ function has two parameters: d_0 and D. While D^f performs data-normalization, d_0^f is the object acceptance threshold concerning feature f: u is attractive according to the $\varphi_f(u)$ term iff $f(u) > d_0^f$.

Finally, we must consider, that the decision based on a single feature f can lead to a *weak classification*, since the buildings and the background may overlap in the f-domain. Therefore, in the *third step* (Sec. 3.4.2), the joint energy term $\varphi(u)$ must be appropriately constructed from the different $\varphi_f(u)$ feature modules.

3.4.1 Feature Models

We begin with `gradient` analysis. Below the edges of a relevant rectangle candidate R_u, we expect the magnitudes of the local gradient vectors (∇g_s) to be high and the orientations to be close to the normal vector ($\mathbf{n}_s$) of the closest rectangle side (Fig. 7). The $f^{\rm gr}(u)$ feature is calculated as:

$$f^{\rm gr}(u) = \frac{1}{\#\tilde{\partial} R_u} \sum_{s \in \tilde{\partial} R_u} \nabla g_s \cdot \mathbf{n}_s \tag{9}$$

where '$\cdot$' denotes scalar product, $\tilde{\partial} R_u$ is the dilated edge mask of rectangle R_u, and $\#\tilde{\partial} R_u$ is the number of pixels in $\tilde{\partial} R_u$. The dilation of the R_u mask outline is necessary to tolerate slightly imperfect edge alignment and minor registration errors between the images. The data-energy term is calculated as: $\varphi_{\rm gr}(u) = \mathcal{Q}(f^{\rm gr}(u), d^{\rm gr}, D^{\rm gr})$.

The calculation of the `roof color` feature is shown in Fig. 9. We expect the image points to have dominantly roof colors inside the building footprint R_u, while the T_u object-neighborhood (see Fig. 9) should contain a majority of background pixels. Hence we calculate the internal $f_{\rm in}^{\rm co}(u)$ and external $f_{\rm ex}^{\rm co}(u)$ filling factors, respectively, as:

$$f_{\rm in}^{\rm co}(u) = \frac{1}{\#R_u} \sum_{s \in R_u} \varrho_{\rm co}(s); \quad f_{\rm ex}^{\rm co}(u) = \frac{1}{\#T_u} \sum_{s \in T_u} \left[1 - \varrho_{\rm co}(s)\right]$$

(a) Object candidate (b) Gradient map (c) Masked gradient map

Fig. 7. Utility of the gradient feature

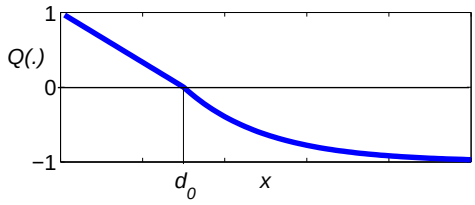


Fig. 8. Plot of the $\mathcal{Q}(x, d_0, D)$ function

Here $\#X$ denotes the area of X in pixels and $\varrho_{\rm co}(s)$ is the color mask value by s. We prescribe that u should be attractive according to the color term if it is attractive both regarding the internal and external subterms. Thus the color energy term is obtained as:

$$\varphi_{\rm co}(u) = \max\left[\mathcal{Q}(f^{\rm co}_{\rm in}(u), d^{\rm co}_{\rm in}, D^{\rm co}_{\rm in}), \mathcal{Q}(f^{\rm co}_{\rm ex}(u), d^{\rm co}_{\rm ex}, D^{\rm co}_{\rm ex})\right]$$

We continue with the description of the `shadow term`. This step is based on the binary shadow mask $\varrho_{\rm sh}(s)$, extracted in Sec. 3.1.2. Using the *shadow direction* vector $\vec{v}_{\rm sh}$ (opposite of the Sun direction vector) we identify the two sides, $\vec{AB}$ and $\vec{BC}$, of the rectangle R_u which are supposed to border on cast shadows, where A, B and C denote the corresponding vertices as shown in Fig. 10. (Note that if $\vec{v}_{\rm sh}$ is parallel to one of the rectangle sides, we have only one shadow-object edge). Then, we check the presence of shadows in parallelograms $(A, A+\varepsilon_{\rm sh}, B+\varepsilon_{\rm sh}, B)$ and $(B, B+\varepsilon_{\rm sh}, C+\varepsilon_{\rm sh}, C)$. Here $\varepsilon_{\rm sh}$ is a scalar so that $||\varepsilon_{\rm sh} \cdot \vec{v}_{\rm sh}||$ approximates the shadow

(a) Red roof (b) Color mask

Fig. 9. Utility of the color roof feature

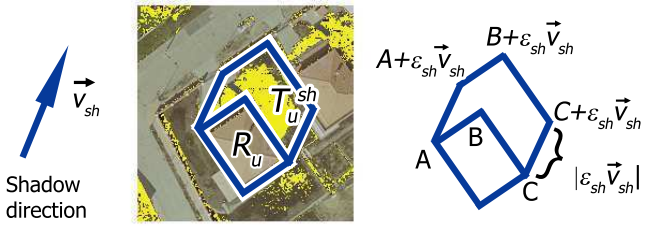
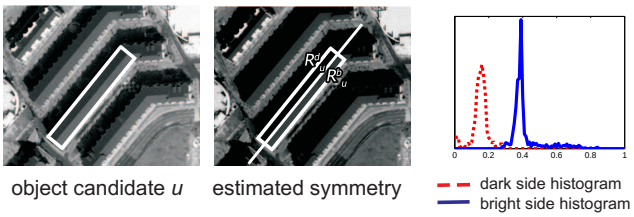


Fig. 10. Utility of the shadow feature



object candidate u estimated symmetry

Fig. 11. Utility of the roof homogeneity feature

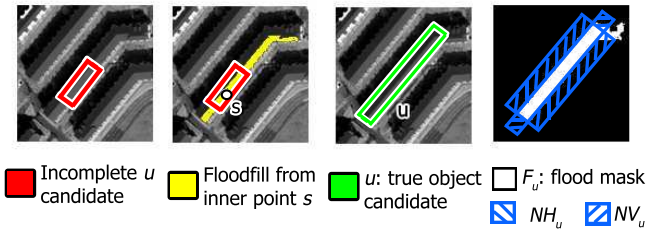


Fig. 12. Floodfill based feature for roof completeness

width of the shortest buildings in the scene. The union of the two parallelograms forms the $T^{\rm sh}_u$ shadow candidate region as shown in Fig. 10. Thereafter, similarly to the color feature, expect low shadow presence $f^{\rm sh}_{\rm ex}(u)$ in the R_u internal and a high one $f^{\rm sh}_{\rm ex}(u)$ in the $T^{\rm sh}_u$ external region:

$$f^{\rm sh}_{\rm in}(u) = \frac{1}{\#R_u}\sum_{s\in R_u}\left[1-\varrho_{\rm sh}(s)\right];\ f^{\rm sh}_{\rm ex}(u) = \frac{1}{\#T^{\rm sh}_u}\sum_{s\in T^{\rm sh}_u}\varrho_{\rm sh}(s)$$

As for the energy term:

$$\varphi_{\rm sh}(u) = \max\left[\mathcal{Q}(f^{\rm sh}_{\rm in}(u), d^{\rm sh}_{\rm in}, D^{\rm sh}_{\rm in}), \mathcal{Q}(f^{\rm sh}_{\rm ex}(u), d^{\rm sh}_{\rm ex}, D^{\rm sh}_{\rm ex})\right]$$

Note that this approach does not require accurate building height information, since we do not penalize it, if shadow blobs of long buildings exceed the $T^{\rm sh}_u$ regions.

The `roof homogeneity` feature can also be exploited at object level. Fig. 11 shows an example of how to describe two-sided roofs. After extracting the symmetry axis of the object candidate u, we can characterize the peakiness of the dark (d) and bright (b) side histograms by calculating their kurtosis $f^{\rm ho}_d(u)$, and $f^{\rm ho}_b(u)$, respectively. Denoting by g_s the gray value of pixel s, and by R^d_u and R^b_u the dark and bright regions of R_u object rectangle, we get:

$$f^{\rm ho}_d(u) = \frac{\sum_{R^d_u} g_s^4}{\left(\sum_{R^d_u} g_s^2\right)^2};\quad f^{\rm ho}_b(u) = \frac{\sum_{R^b_u} g_s^4}{\left(\sum_{R^b_u} g_s^2\right)^2} \qquad (10)$$

If the roof parts are homogeneous, the $f^{\rm ho}_d(u)$ and $f^{\rm ho}_b(u)$ kurtosis values should be high. However, as shown in Fig. 12, the homogeneity feature may have false maxima for incomplete roofs, since parts of a homogeneous roof are homogeneous as well. Therefore we characterize roof completeness in the following way. We derive the F_u floodfill mask of u, which contains the pixels reached by floodfill propagations from the internal points of R_u. If the homogeneous roof is complete, F_u must have low intersection with the NH_u, resp. NV_u, 'horizontal', and 'vertical', neighborhood regions of R_u (see Fig. 12). Finally, the $\varphi_{\rm ho}(u)$ energy term can be

constructed from the kurtosis and completeness descriptors in a similar manner to the previous attributes.

3.4.2 Feature integration

Usually, the individual features are in themselves inappropriate for modeling complex scenes, which is illustrated in Fig. 13. For the Ground Truth (GT) buildings (see Fig. 13(b)), we can follow here the effectiveness of the gradient (Fig. 13(c),(d)), shadow (Fig. 13(e),(f)) and color (Fig. 13(g),(h))) descriptors, respectively. The gradient and shadow maps are considerably noisy in the right upper image part, however, the roofs can be detected here fairly by extracting image regions with high a* color component values in CIE L*a*b* color space representation. Conversely, '*non-red*' buildings in the bottom-left regions can be efficiently detected by edge and shadow features.

To answer the challenges of such object or data heterogeneity problems, the proposed framework enables flexible *feature integration* depending on the available image inputs. From the feature primitive terms introduced in Sec. 3.4, first we construct building prototypes. For each prototype we can prescribe the fulfillment of one or many feature constraints whose φ_f-subterms are connected with the $\max$ operator in the joint energy term of the prototype (logical AND in the negative log-likehood domain).

Additionally, several building prototypes can be detected simultaneously in a given image pair, if the prototype-energies are joined with the $\min$ (logical OR) operator. Thus the final object energy term is derived by a logical function, which expresses some prior knowledge about the image and the scene, and it is chosen on a case-by-case basis. For example, in the BUDAPEST pair we use two prototypes: the first one prescribes the edge and shadow constraints, the second one the roof color, thus the joint energy is calculated as:

$$\varphi(u) = \min\big\{\max\{\varphi_{\mathrm{gr}}(u), \varphi_{\mathrm{sh}}(u)\}, \varphi_{\mathrm{co}}(u)\big\}. \qquad (11)$$

Similarly, for the BEIJING images (see Fig. 26, bottom) we use gradient (φ_{gr}) & shadow (φ_{sh}) and homogeneity (φ_{ho}) & shadow (φ_{sh}) prototypes.

4 CONFIGURATION MODEL AND OPTIMIZATION

In this section we transform the building change detection task into an energy minimization problem. Following our definitions from Sec. 2, the u building segment candidates (i.e. *objects*) live in a bounded parameter space $\mathcal{H}$. Since we aim to extract building populations from the images, we need to propose a configuration space Ω, which is able to deal with an unknown number of objects:

$$\Omega = \bigcup_{n=0}^{\infty} \Omega_n, \qquad \Omega_n = \big\{\{u_1, \ldots, u_n\} \subset \mathcal{H}^n\big\} \qquad (12)$$

Hereafter we will use the notation $\omega \in \Omega$ for an arbitrary object configuration, thus $\omega = \emptyset$, or $\omega = \{u_1, \ldots, u_n\}$ for an $n \in \{1, 2, \ldots\}$ and $u_i \in \mathcal{H} : \forall i \in \{1, 2, \ldots, n\}$.

4.1 Configuration Energy

The Marked Point Process framework enables to characterize whole populations instead of individual objects, through exploiting information from entity interactions. Following the classical Markovian approach, each object may only affect its *neighbours* directly. This property limits the number of interactions in the population and results in a compact description of the global scene, which can be analyzed efficiently. To realize the Markov-property, one should define first a $\sim$ neighborhood relation between the objects in $\mathcal{H}$. In our model, we say that $u \sim v$ if their rectangles R_u and R_v intersect.

Let us denote by $\mathcal{D}$ the union of all image features derived from the input data. For characterizing a given ω object population considering $\mathcal{D}$, we introduce a non-homogenous data-dependent Gibbs distribution on the configuration space:

$$P_{\mathcal{D}}(\omega) = \frac{1}{Z} \cdot \exp\Big[-\Phi_{\mathcal{D}}(\omega)\Big] \qquad (13)$$

with a Z normalizing constant: $Z = \sum_{\omega \in \Omega} \exp\big[-\Phi_{\mathcal{D}}(\omega)\big]$, and $\Phi_{\mathcal{D}}(\omega)$ configuration energy:

$$\Phi_{\mathcal{D}}(\omega) = \sum_{u \in \omega} A_{\mathcal{D}}(u) + \gamma \cdot \sum_{\substack{u,v \in \omega \\ u \sim v}} I(u, v) \qquad (14)$$

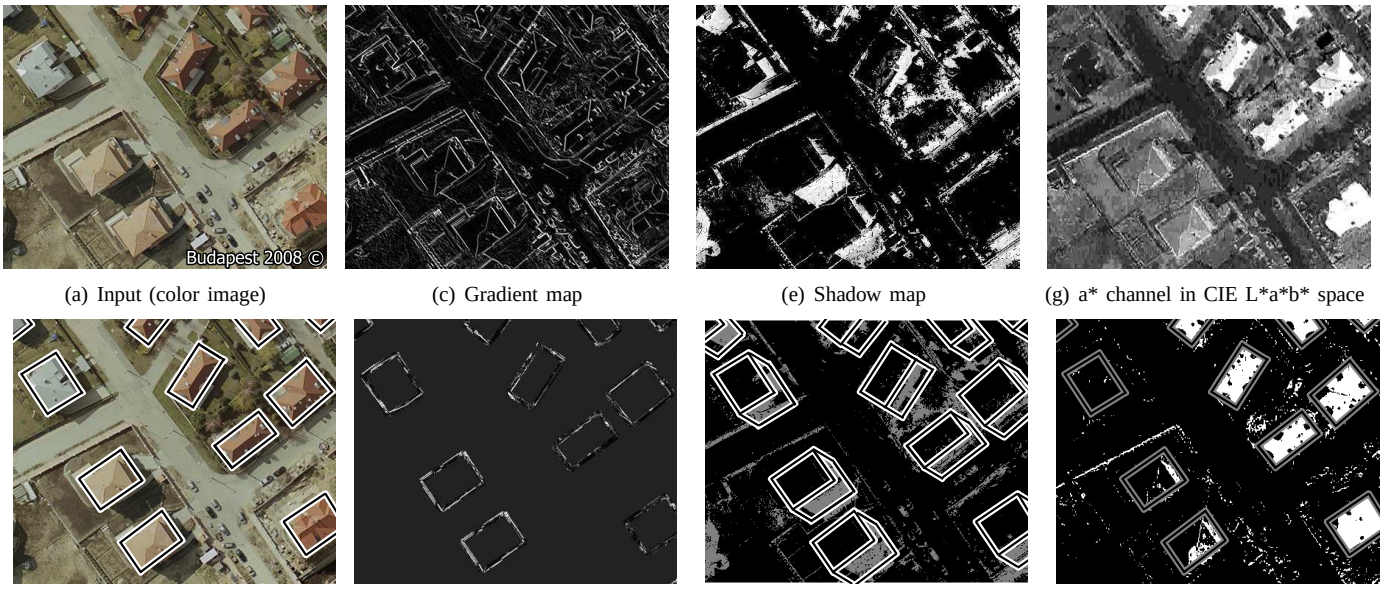
Here $A_{\mathcal{D}}(u) \in [-1, 1]$ and $I(u, v) \in [0, 1]$ are the data dependent unary and the prior interaction potentials, respectively, and $\gamma > 0$ is a weighting factor between the two energy terms. Thus the Maximum Likelihood (ML) configuration estimate according to $P_{\mathcal{D}}(\omega)$ can be calculated as $\omega_{\mathrm{ML}} = \mathrm{argmin}_{\omega \in \Omega}\big[\Phi_{\mathcal{D}}(\omega)\big]$.

Unary potentials characterize a given building segment candidate $u = \{c_x, c_y, e_L, e_l, \theta, \xi\}$ as a function of the local image data in both images, but independently of other objects of the population. This term encapsulates the building energies $\varphi^{(1)}(u)$ and $\varphi^{(2)}(u)$ extracted from the 1^{st}, resp. 2^{nd}, image (Sec. 3.4) and the low level similarity information between the two time layers which is described by the $\varrho_{\mathrm{ch}}(.)$ change mask (Sec. 3.2).

We remind the reader that our approach marks each building segment u with an image index flag from the set $\{1, 2, *\}$, depending on that u appears in one [$\xi(u) \in \{1, 2\}$] or both [$\xi(u) = *$] of the input images. In this way, the classification of the building segment u is straightforward: u is *unchanged* iff $\xi(u) = *$; *new* iff $\xi(u) = 2$ and $\nexists v \in \omega : \{\xi(v) = 1, u$ and v overlap$\}$; and *demolished* iff $\xi(u) = 1$ and $\nexists v \in \omega : \{\xi(v) = 2, u$ and v overlap$\}$. Modified buildings are considered as two objects u_1 and u_2, so that $\xi(u_1) = 1$, $\xi(u_2) = 2$.

The following *soft constraints* are considered by the potential terms in the various cases:

- unchanged building u: we expect low object energies in both images, and penalize textural differences (i.e. pixels with $\varrho_{\mathrm{ch}}(s) = 1$) under its footprint R_u.
- demolished or modified building in the first image: we expect low $\varphi^{(1)}(u)$, and $\varphi^{(2)}(u)$ is indifferent. We penalize high similarity under the footprint.
- new or modified building in the second image: we expect low $\varphi^{(2)}(u)$, and $\varphi^{(1)}(u)$ is indifferent. We penalize high similarity under the footprint.



(a) Input (color image) (c) Gradient map (e) Shadow map (g) a* channel in CIE L*a*b* space

(b) Ground Truth (GT) (d) Gradient feature (f) Shadow feature (h) Color mask

Fig. 13. Illustration of the feature maps in the BUDAPEST 2008 image. Gradient and shadow features are relevant in the left-bottom regions, while the color descriptor is efficient in the top-right image parts. In image (d), the gradient feature is shown under the GT object borders and the background color is equal to the average gradient value.

Consequently, using the $\mathbb{I}_{[.]} \in \{0,1\}$ indicator function for an event noted in the subscript $[.]$, the $A_{\mathcal{D}}(u)$ potential is calculated as:

$$A_{\mathcal{D}}(u) = \mathbb{I}_{[\xi(u)\in\{1,*\}]} \cdot \varphi^{(1)}(u) + \mathbb{I}_{[\xi(u)\in\{2,*\}]} \cdot \varphi^{(2)}(u) + \\ + \mathbb{I}_{[\xi(u)=*]} \cdot \frac{1}{\#R_u} \sum_{s\in R_u} \varrho_{\mathrm{ch}}(s) + \\ + \mathbb{I}_{[\xi(u)\in\{1,2\}]} \cdot \frac{1}{\#R_u} \sum_{s\in R_u} \big(1 - \varrho_{\mathrm{ch}}(s)\big) \qquad (15)$$

On the other hand, *interaction* potentials realize prior geometrical constraints: they penalize intersection between different object rectangles sharing the time layer (see Fig. 14):

$$I(u,v) = \mathbb{I}_{[\xi(u)\simeq\xi(v)]} \cdot \frac{\#(R_u \cap R_v)}{\#(R_u \cup R_v)} \qquad (16)$$

where $\xi(u) \simeq \xi(v)$ relation holds iff $\xi(u) = \xi(v)$, or $\xi(u) = *$, or $\xi(v) = *$. Since $\forall u, v : I(u,v) \geq 0$, the optimal population should exclusively consist of objects with negative data terms (i.e. attractive objects): if $A_{\mathcal{D}}(u) > 0$, removing u from the configuration results in a lower $\Phi_{\mathcal{D}}(\omega)$ global energy (14). Note also that according to eq. (14), the interaction term plays a crucial role by penalizing multiple attractive objects in the same or strongly overlapping positions.

Note that in the introduced probabilistic model, it is also possible to involve additional prior knowledge about the layout of settlements, by adding further prior terms to the global energy function $\Phi_{\mathcal{D}}(\omega)$. For example, in a town, buildings are usually aligned, hence, we can use the geometric interaction terms of [6], where the *alignment constraint* favors small angle difference and low distance between appropriately matched corners of neighboring segments, while the *paving constraint* favors parallel rectangles that are located side by side inducing clean arrangements of buildings.

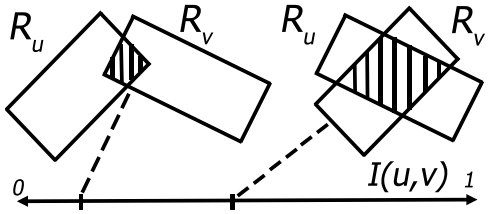


Fig. 14. Calculation of the $I(u,v)$ interaction potentials: intersections of rectangles are denoted by striped areas

4.2 Bi-layer Multiple Birth and Death Optimization

By fixing the $A_{\mathcal{D}}(u)$ and $I(u,v)$ potential terms, the $\Phi_{\mathcal{D}}(\omega)$ configuration energy is completely defined, and the optimal ω_{ML} building population can be obtained by minimizing eq. (14). For this purpose, we have developed the *bi-layer Multiple Birth and Death* (bMBD) algorithm, the main steps can be followed in Fig. 15. The bMBD method extends the conventional MBD technique by handling two time layers, thus it encapsulates change and object information simultaneously. Pairs of consecutive birth and death processes are iterated until convergence is obtained in the global configuration. In the *birth* step, multiple object candidates are generated randomly according to the birth maps $P_b^{(i)}(s)$, and as a further novelty, also considering the change probabilities $P_{\mathrm{ch}}(s)$ with the expected parameter maps $\mu_\theta^{(i)}(s)$, $\mu_L^{(i)}(s)$ and $\mu_l^{(i)}(s)$ $i \in \{1,2\}$. The *death* process attempts to eliminate the inappropriate objects based on the global configuration energy.

5 MULTI-SCALE GENERALIZATION

In Sec. 2, we have used a single size hypothesis for all buildings in the image. However, in scenes where the sizes

Bi-layer Multiple Birth and Death (bMBD) algorithm

1) Initialization: calculate the $P_b^{(i)}(s)$, $P_{\mathrm{ch}}(s)$, $\mu_\theta^{(i)}(s)$, $\mu_L^{(i)}(s)$ and $\mu_l^{(i)}(s)$ ($i \in \{1,2\}$) birth maps, and start with an empty population $\omega = \emptyset$.
2) Main program: initialize the inverse temperature parameter $\beta = \beta_0$ and the discretization step $\delta = \delta_0$ and alternate birth and death steps:
   - *Birth step*: for each pixel $s \in S$, if there is no object with center s in the current configuration ω, pick up $\xi \in \{1, 2, *\}$ randomly, let be

$$\widehat{P}_b = \begin{cases} P_{\mathrm{ch}}(s) \cdot P_b^{(\xi)}(s) & \text{if } \xi \in \{1,2\} \\ \big(1 - P_{\mathrm{ch}}(s)\big) \cdot \max\{P_b^{(1)}(s), P_b^{(2)}(s)\} & \text{if } \xi = * \end{cases}$$

     and execute the following birth process with probability $\delta \widehat{P}_b$:
     – generate a new object u with center s and image index ξ
     – set the $e_L(u)$ and $e_l(u)$ side length parameters as follows:
       * with a probability $\widehat{P}_b^h(s)/\widehat{P}_b$: set the parameters according to $\eta\big(., \mu_L^{(\xi)}(s), \sigma_L\big)$ resp. $\eta\big(., \mu_l^{(\xi)}(s), \sigma_l\big)$ Gaussian distributions as explained in Sec. 3.1.3. (Notes: (i) If the homogeneity feature is ignored, $\widehat{P}_b^h(s)$ is considered as a constant zero map. (ii) If $\xi = *$, we choose between $\mu_{L/l}^{(1)}$ and $\mu_{L/l}^{(2)}$ randomly.)
       * otherwise: set the parameters randomly between prescribed maximal and minimal side lengths, following a uniform distribution
     – set the orientation $\theta(u)$ following the $\eta(., \mu_\theta^{(\xi)}(s), \sigma_\theta)$ Gaussian distribution as shown in Sec. 3.1
     – add u to the current configuration ω
   - *Death step*: Consider the configuration of objects $\omega = \{u_1, \ldots, u_n\}$ and sort it from the highest to the lowest value of $A_{\mathcal{D}}(u)$. For each object u taken in this order, compute $\Delta\Phi_\omega(u) = \Phi_{\mathcal{D}}(\omega/\{u\}) - \Phi_{\mathcal{D}}(\omega)$, derive the *death rate* as follows:

$$d_\omega(u) = \frac{\delta a_\omega(u)}{1 + \delta a_\omega(u)}, \quad \text{with} \quad a_\omega(u) = e^{-\beta \cdot \Delta\Phi_\omega(u)}$$

     and remove u from ω with probability $d_\omega(u)$. Note that according to eq. (14), $\Delta\Phi_\omega(u)$ depends only on u and its neighbours in ω, thus $d_\omega(u)$ can be calculated locally without computing the global configuration energies $\Phi_{\mathcal{D}}(\omega/\{u\})$ and $\Phi_{\mathcal{D}}(\omega)$.
   - *Convergence test*: if the process has not converged, increase the inverse temperature β and decrease the discretization step δ by a geometric scheme and go back to the birth step. Convergence is obtained when all the objects added during the birth step, and only these ones, have been killed during the death step.

Fig. 15. Pseudo code of the bi-layer Multiple Birth and Death (bMBD) algorithm

TABLE 1
Main properties of the test data sets, and applicable features from Sec. 3 ($\surd$=Yes, $\times$=No).

Data Set	Type	Source	Ch. det†	Usable features			
				co	sh	gr	ho
BUDAPEST	Aerial	City Council	$\surd$	$\surd$	$\surd$	$\surd$	$\times$
ABIDJAN	Satellite	Ikonos	$\surd$	$\times$	$\times$	$\surd$	$\surd$
BEIJING	Satellite	QuickBird	$\surd$	$\times$	$\surd$	$\surd$	$\surd$
SZADA	Aerial	FÖMI‡	$\surd$	$\surd$	$\times$	$\surd$	$\times$
C. D'AZUR	Satellite	Google Earth	$\times$	$\surd$	$\surd$	$\surd$	$\times$
BODENSEE	Satellite	Google Earth	$\times$	$\surd$	$\surd$	$\surd$	$\times$
NORMANDY	Satellite	Google Earth	$\times$	$\surd$	$\surd$	$\surd$	$\times$
MANCHESTER	Satellite	Google Earth	$\times$	$\surd$	$\surd$	$\surd$	$\times$

†indicate if multiple time layers are available for change detection
‡Hungarian Inst. of Geodesy, Cartography and Remote Sensing

of the footprints are notably diverse (see Fig. 17 and 18), this simplification may prove to be inefficient. Considering multiple scales concerns the low level feature extraction part and the *Birth* step of the bMBD process particularly (see Fig. 15), since the object level features (Sec. 3.4) calculated in the *Death* step are normalized either with the object area or with the perimeter. In this section, we demonstrate a generalization of the method through providing a multi-scale extension of the Gradient Orientation Density Function (GODF), $\lambda_s(.)$, and introducing further modifications which are necessary in the proposed bMBD algorithm. We also note that most of the other low level features can be handled in a similar manner.

Let us assume that we have J different building size hypothesizes: $\forall u : u \in \cup_{j=1}^{J} \Upsilon_j$ where $u \in \Upsilon_j$ iff $e_L(u) \in [e_{L,j}^{\min}, e_{L,j}^{\max}]$ and $e_l(u) \in [e_{l,j}^{\min}, e_{l,j}^{\max}]$.

We remind the reader that the $\lambda_s(.)$ feature has been calculated over a rectangular image region $W_l(s)$ around pixel s, where the l window side length has been set according to the estimated average object size. In the multi-scale extension, we calculate the local GODF for J different window sizes ($l = l_1, \ldots, l_J$) corresponding to the J size hypotheses. Fig. 16 demonstrates this process with $J = 3$, $l_1 = 14$, $l_2 = 24$ and $l_3 = 40$ parameter settings. At each scale $j = 1 \ldots J$, we calculate the $P_b^{\mathrm{gr}}(s, j)$ birth probabilities and $\mu_\theta(s, j)$ mean orientation estimates separately, and get the final gradient-birth-map as $P_b^{\mathrm{gr}}(s) = \max_j P_b^{\mathrm{gr}}(s, j)$.

A minor modification should also be inserted into the birth map summarization step: we store at each pixel s the

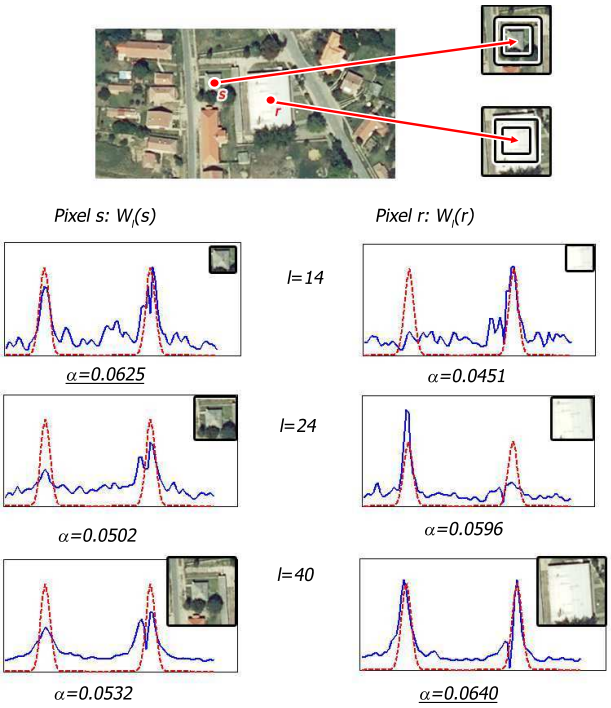


Fig. 16. Multi-scale investigations: demonstrating the dependency of gradient orientation histograms on the l window size. Maximal α feature can be obtained with $l = 14$ for the small building (see on the left) and with $l = 40$ for the large building (see on the right)

(a) Input image (b) Detection

Fig. 17. Detection results in case of significantly different building scales, using the features from Fig. 16

dominant feature term, $\chi(s) = \text{argmax}_{\chi\in\{\text{gr,co,sh,ho}\}} P_b^{\chi}(s)$. This indicator is utilized by the modified *Birth* process. If we decide to generate a new object u at pixel s based on $\widehat{P}_b(s)$, we select its scale $j(u)$ randomly, so that the probability of choosing scale j is $P_b^{\chi(s)}(s,j)/\sum_{i=1}^{J} P_b^{\chi(s)}(s,i)$. Then we set the orientation $\theta(u)$ following the $\eta(.,\mu_\theta(s,j(u)),\sigma_\theta)$ distribution, and the side lengths $e_L(u)$ and $e_l(u)$ according to uniform distributions around the expected length values at scale $j(u)$.

6 Parameter Settings

We can divide the parameters of the proposed mMPP method into three groups corresponding to the *prior model*, *data model* and the *bMBD optimization*.

The *prior model* parameters, such as the number of the examined scales (J in Sec. 5), l (or l_j) window sizes for GODF calculation (see Sec. 3.1.1) and maximal/minimal rectangle side lengths at the difference scales, depend on image resolution and expected object dimensions. They are set based on sample objects. We used a constant $\gamma = 2$ weight between the data term and the overlapping coefficient in (eq. 14).

The parameters of the *data model* are estimated based on training image regions containing *Ground Truth* building segments $\{u_1^{\text{gt}}, u_2^{\text{gt}}, \ldots, u_n^{\text{gt}}\}$. Consider an arbitrary $f(u)$ feature from the feature library (e.g. $f^{\text{gr}}(u)$ gradient descriptor). We remind the reader that each $f(u)$ of our model is a noisy quality measure and the corresponding energy term is obtained as $\varphi_f(u) = \mathcal{Q}(f(u), d_0^f, D^f)$ (see Sec. 3.4). Here we set the normalizing constant as $D^f = \max_j f(u_j^{\text{gt}}) - \min_j f(u_j^{\text{gt}})$. Exploiting that the $\mathcal{Q}$ transfer function is monotonously decreasing with a sole root $f(u) = d_0^f$, object u is attractive in image i (i.e. $\varphi_f^{(i)}(u) < 0$) iff $f(u) > d_0^f$. Consequently, increasing d_0^f may decrease the false alarm rate and increase the missing alarms corresponding to the selected feature. Since in the proposed model we can simultaneously utilize several building prototypes, our strategy for setting d_0^f is to minimize the false alarms for each prototype, and eliminate the missing buildings using further feature tuples.

Finally, regarding the *relaxation* parameters, we followed the guidelines provided in [22], and used $\delta_0 = 20000$, $\beta_0 = 50$, and geometric cooling factors $1/0.96$.

(a) Input image (b) Detection

Fig. 18. Detection results in a densely built-in part of the ABIDJAN image set

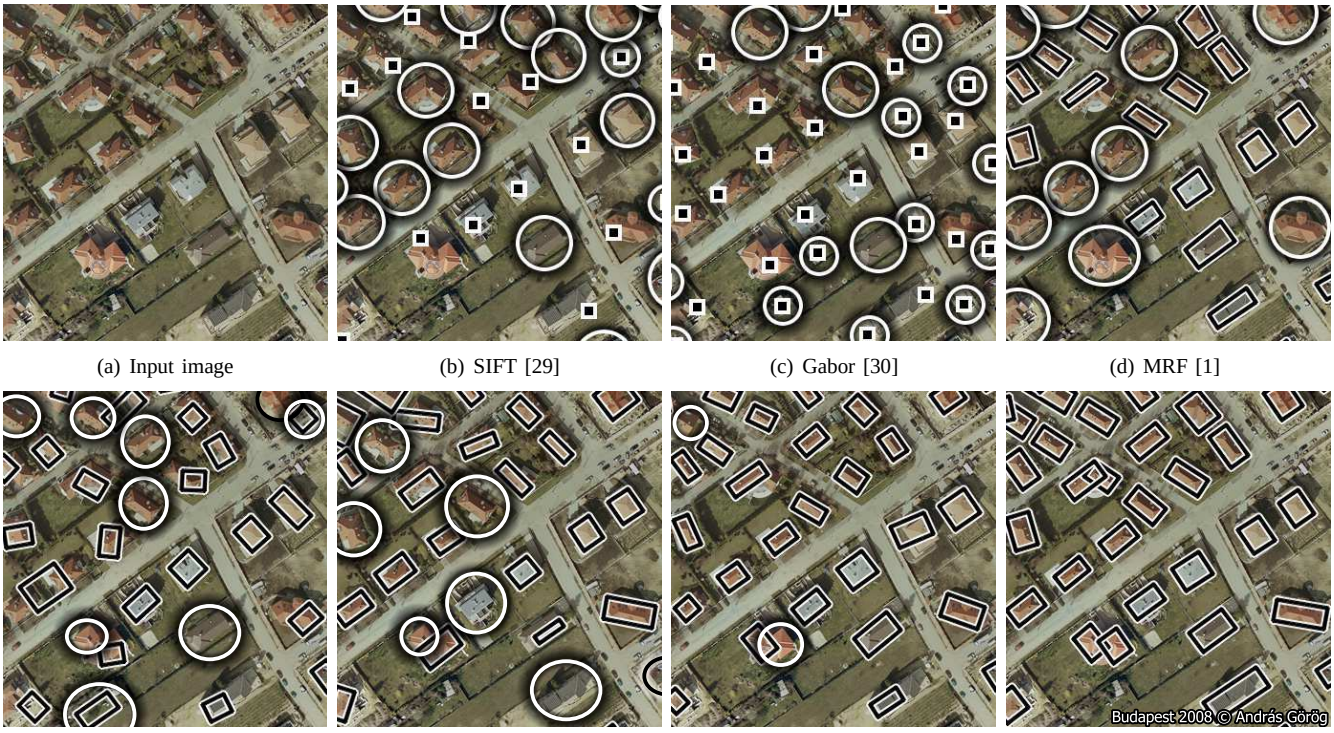
Edge Verification method [3]

- Building Candidate Region (BCR) extraction with a morphological approach – shadow and roof color information exploited *solely* in preprocessing
- Canny edge detection inside the BCRs
- Roof corner estimation by detection of perpendicular edge-junctions in the BCRs
- Rectangle fitting for the BCR-edge map around each corner candidate
- Hypothesis acceptance/rejection

Segment-Merge method [17]

- Building segment estimation by seeded region growing
- Region merging and shadow evidence verification
- Filtering based on geometric and photometric features
- Polygon approximation of the building blocks

Fig. 19. Main steps of the Edge Verification [3] and Segment-Merge [17] methods used for comparison



(a) Input image (b) SIFT [29] (c) Gabor [30] (d) MRF [1]

(e) Edge Verification [3] (f) Segment-Merge [17] (g) Proposed MPP (h) Ground Truth

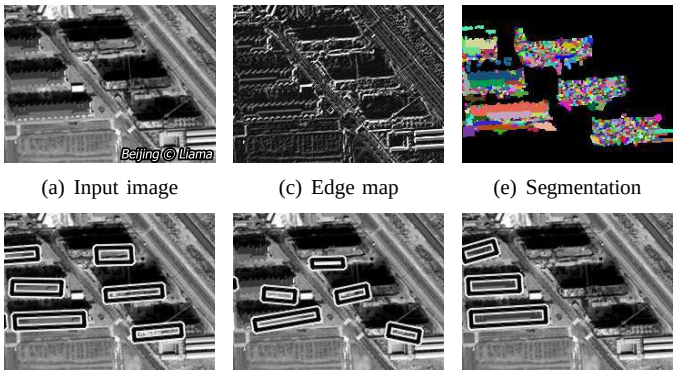
Fig. 20. Evaluation of the single view building model. Comparing the proposed MPP model to the SIFT [29], Gabor [30], MRF [1], Edge Verification (EV) [3], Segment-Merge (SM) [17] methods, and to the Ground Truth. Circles denote completely missing or false objects. SIFT and Gabor only extract building centers.

7 EXPERIMENTS

The goal of this section is to validate the three key developments of the paper and compare them to the state of the art: (i) the proposed multiple feature based building appearance model, (ii) the joint object-change modeling framework and (iii) the non-homogeneous object birth process based on low level features.

We have evaluated our method using eight significantly different data sets whose main properties are summarized in Table 1. Four image collections contain multitemporal aerial or satellite photos from the monitored regions, which enables testing both the building extraction and the change detection abilities of the proposed mMPP model. The remaining four data sets contain standalone satellite images acquired from Google Earth, which are only exploited in the evaluation of the building appearance model (Sec. 7.1). To guarantee the heterogeneity of the test sets, we have chosen completely different geographical regions as listed in Table 1. We collected samples from densely populated suburban areas, and built a manually annotated database for validation. For parameter settings, we have chosen in each data set 2-8 buildings ($\approx 5\%$) as training data, while the remaining Ground Truth labels have only been used to validate the detection results. Qualitative results are shown in Fig. 17, 18, 20, 21, 25 and 26.

We perform quantitative evaluation both at object and pixel levels. On one hand, we measure how many buildings are recognized or classified incorrectly in the different test sets, by counting the missing and falsely detected objects (MO and FO, respectively), and the missing and false change alarms (MC, FC). On the other hand, we also investigate how accurate the extracted object outlines are: we compare the resulting building footprint masks to the Ground Truth mask, and calculate the Precision (Pr) and Recall (Rc) values of the pixel level detection. Finally, the F-score (harmonic mean of Pr and Rc) can be given both at object and at pixel levels.



(a) Input image (c) Edge map (e) Segmentation

(b) Proposed MPP (d) EV method (f) SM method

Fig. 21. Limitations of the EV and SM methods: compared to the proposed model (im. (b)), weak edge map (c) results in weak EV matching (d); while textured buildings on the right do not appear as homogenous blobs in the floodfill map (e), and are ignored by SM detection (f)

TABLE 2
Numerical object level and pixel level comparison of the SIFT, Gabor, EV, SM and the proposed methods (MPP) on each test data set (best results in each row are typeset by bold.)

Data Set		Object level performance										Pixel level performance					
		SIFT [29]		Gabor [30]		EV [3]		SM [17]		**Prop. MPP**		EV [3]		SM [17]		**Prop. MPP**	
Name	#obj.	MO	FO	MO	FO	MO	FO	MO	FO	MO	FO	Pr	Rc	Pr	Rc	Pr	Rc
BUDAPEST	41	20	10	8	17	11	5	9	**1**	**2**	4	0.73	0.46	**0.84**	0.61	0.82	**0.71**
ABIDJAN	21	8	5	**0**	1	2	**0**	2	1	1	**0**	**0.91**	0.73	0.84	**0.79**	0.83	0.74
BEIJING	17	7	2	9	8	2	3	4	2	**1**	**0**	0.59	0.26	0.71	**0.72**	**0.93**	0.71
SZADA	57	17	26	17	23	10	18	11	5	**4**	**1**	0.61	0.62	0.79	0.71	**0.93**	**0.75**
CÔTE D'AZUR	123	55	9	12	24	14	20	20	25	**5**	**4**	0.73	0.51	0.75	0.61	**0.83**	**0.69**
BODENSEE	80	34	9	32	8	11	13	18	15	**7**	**6**	0.56	0.30	0.59	0.41	**0.73**	**0.51**
NORMANDY	152	69	14	24	14	**18**	32	30	58	**18**	**1**	0.60	0.32	0.62	0.55	**0.78**	**0.60**
MANCHESTER	171	NA	NA	53	85	46	17	53	42	**19**	**6**	0.64	0.38	0.60	0.56	**0.86**	**0.63**
Overall F-score*		0.663		0.799		0.842		0.798		**0.944**		0.537		0.668		**0.743**	

* MANCHESTER is ignored from the summarization due to weak performance with most of the methods

7.1 Building Segment Description

Although the proposed model handles multiple time layers simultaneously, the building description module introduced in Sec. 3.4 works on single image inputs (birth maps and object energies are calculated in the two images independently). In this subsection, we evaluate solely the object recognition part, therefore we use temporarily a simplified data term $A_{\mathcal{D}}(u) = \varphi^{(2)}(u)$, i.e. we detect buildings only in the second image independently of the first one.

In this section, we present numerical and qualitative comparison results versus single-view building detection techniques from the state-of-the-art, which were briefly introduced in Sec. 1.4. On one hand, we have evaluated three recent methods in collaboration with their authors: MRF [1], SIFT [29], and Gabor [30]. Source codes of SIFT and Gabor have been provided us for the experiments, while the authors of MRF have tested their model with our image sets. In addition, we have also implemented two methodologically orthogonal methods: the Edge Verification (EV) technique [3] and the Segment-Merge (SM) model [17], their main steps are listed in Fig 19. Since EV and SM use similar image features (gradient, shadow, color, homogeneity) to our framework, by considering them in the comparison, we can focus purely on validating the model structures instead of special input-dependent descriptors. Sample output images of the reference methods can be found in Fig. 20 and 21.

Quantitative evaluation on the database was performed with SIFT, Gabor, EV, SM and the proposed MPP models, results are shown in Table 2. Since SIFT and Gabor extract the building centers instead of estimating the outline, they are only involved in the object level comparison. Numerical results confirm that the proposed model surpasses all references with 10-26% at object level and with 5-18% at pixel level.

Table 3 lists the computational time requirements of the test images with the different methods. From this viewpoint, the Gabor technique is dominantly the most efficient, since its Matlab version outperforms the other C^{++}coded methods. On the other hand we can observe that the proposed MPP model is competitive with most reference techniques regarding the average running time as well. Note that as Table 3 demonstrates, the computational complexity for the different images depends in parallel on various factors, such as image size, dominance of the color map, and diversity of the building side length values.

Apart from the low complexity, a significant advantage of the Gabor model [30] is that it can deal with various images by changing only a single scale parameter. Conversely, our method uses 1-2 free parameters for each feature, thus it is more dependent on the training set. However, while the *Gabor* algorithm is successful in detecting compact shaped buildings, it faces difficulties with long segments (see BEIJING). Some additional problems appear considering dark buildings (BODENSEE), where the gradient directions point away from the building center [30]. Both the *Gabor* and *SIFT* methods have better performance on panchromatic satellite images, while in aerial photos false positives appear due to many redundant local features extracted in the background. In general, buildings in rural regions are efficiently detected with these models, but in densely populated areas, the false alarm rate increases. On the contrary, inserting various building hypotheses into our MPP framework is straightforward, thus it can efficiently deal both with high contrast satellite images and aerial photos where color is a more dominant feature.

Since our datasets do not contain any metadata about intrinsic and extrinsic camera parameters, several benefits of [1] cannot be exploited. Most of the observed artifacts of [1] (see Fig. 20(d)) derive from the sensitivity of the edge detection algorithm, which is alleviated by using a strong assumption: the detected buildings should have a uniform height. Resolution of the images affects strongly the quality of the results. The authors confirmed that the minimal resolution, at which the method is able to operate is only fulfilled in the HR versions of the Budapest and Côte d'Azur images. They reported fundamental problems with the remaining images, partially due to small size, missing color information and low quality of the edges. Buildings with irregular shapes are not detected because they do not fulfill a second assumption, that rooftops are polygons with pairwise parallel sides.

We continue the discussion with the EV and SM reference methods. Both of them follow the deterministic hypothesis generation-acceptance scheme, where buildings ignored by the

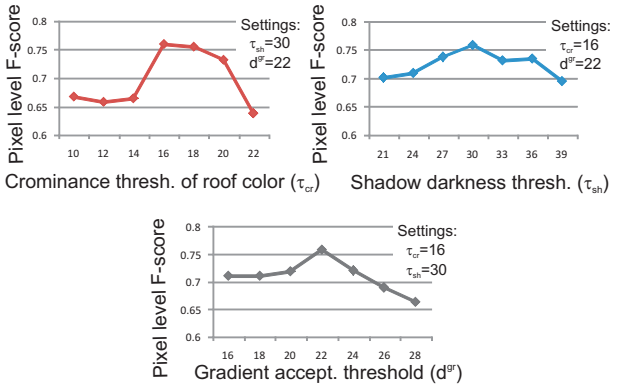


Fig. 22. Pixel level performance (F-scoore) in case of different parameter settings

TABLE 3
Computational time of the different w.r.t. image sizes (in kPixels)

Data Set	Size (kPix)	Computational time (seconds)				
		SIFT	Gabor	EV	SM	**MPP**
BUDAPEST†	280	197.3	**14.5**	120.4	18.4	25.2
ABIDJAN	148	110.1	7.1	12.0	**6.3**	10.7
BEIJING	515	391.1	37.2	155.5	**12.9**	52.2
SZADA	1472	200.9	49.8	**30.3**	89.1	31.5
C. D'AZUR	723	416.0	57.7	324.3	**47.2**	68.5
BODENSEE	536	90.0	**27.9**	30.4	35.1	66.2
NORMANDY	1116	236.7	67.2	109.6	72.3	**46.3**
MANCHESTER	1073	NA	137.1	132.1	**54.0**	65.9
Average	733	234.6	**37.4**	111.8	40.2	42.9
Implementation language:		Matlab	Matlab	C++	C++	C++

†Test of [1] with full resolution (1068kPix) needed 45 *minutes*

hypothesis generator modules appear automatically as missing objects. On the contrary, our proposed model uses a stochastic birth process, where the $P_b(s)$ maps control the frequency of local object propositions only. This approach is more robust, since even in the image regions with erroneously low $P_b(s)$ values, the objects are not completely absent; they merely appear later during the bMBD iterations or the silhouettes may be slightly inaccurate.

As in Sec. 1.4, the sequential EV technique is sensitive to the quality of the individual feature maps used in the consecutive algorithmic steps. Similarly to [1], the edge mask exploited in object proposition and verification can be strongly corrupted or misleading in cases of low contrast, textured background or large noise (Fig. 20(e)). In addition, examples in Fig. 21(c) and (d) show that curved or scalloped roof edges do not fit appropriately to the straight sides of the model rectangles and result in poor detection on the left part of the image. For the same reasons, we must expect similar artifacts by using other edge-critical methods [18], [31].

As for the region based approach, Fig. 20(f) and Fig. 21(e) - (f) illustrate that the SM technique fails in detecting roofs which are inhomogeneous both in color and texture. On the other hand, building-like homogeneous blobs may result in false positive objects, while low contrast buildings can be merged with the background and missed during the segmentation process. These problems can also appear using similar methods [33], [34].

In summary, the tests confirm that the proposed model surpasses the reference techniques, particularly due to its two key properties: the stochastic object generation process and the parallel utilization of multiple features in the building description module.

Finally, we have tested the sensitivity of the proposed model against the parameters of various feature extraction steps. Fig. 22 shows the pixel level F-scores of detection on the BUDAPEST image, where we perturbated the chrominance threshold ($\tau_{\rm cr}$) of roof color filtering, the shadow darkness threshold ($\tau_{\rm sh}$) and the gradient acceptance threshold ($d^{\rm gr}$) with maximum $\pm 30\%$ around the optimal value. Results show that the performance varies around 10% in these parameter domains, most significant is the dependence on $\tau_{\rm cr}$.

7.2 Joint Object-Change Model

After testing the introduced building detector module in single images, we continue with the validation of the proposed joint object-change classification framework. The mMPP model evaluates a given building segment candidate by simultaneously considering its bi-temporal $\varphi^{(1)}(u)$ and $\varphi^{(2)}(u)$ object energies and the low level change information under the footprint. This approach is compared to the conventional Post Detection Comparison (PDC) [13] technique, where the buildings are separately extracted from the two image layers, and the change information is a posteriori estimated through comparing the location, geometry and spectral characteristics of the detected objects. In the latter case the object-change decision is sequential, thus less information can be exploited by the individual object extraction and change classification steps, respectively. Table 4 confirms, that the PDC method causes more false change alarms than mMPP.

To understand the reasons for the differences between PDC and mMPP, a few illustrative examples are shown in Fig. 25. First, the layer-by-layer detector has missed two object candidates: one in the top of the (a) image (edges are partially hidden by the trees), and one in the bottom-left corner of the (b) image (low contrast). These errors result automatically in false changes by using the PDC approach. However, the joint mMPP model produces appropriate detection results (images (c) and (d)), exploiting that both imperceptible buildings are in *certainly unchanged* image parts according to the low level $\varrho_{\rm ch}-$change feature, meanwhile the given objects have been correctly and confidently detected in the other images. On the other hand, false objects appearing in the background regions in PDC have been eliminated by the mMPP model, exploiting that the corresponding local similarity is high again, but the *'twin'* object cannot be found at the other time instance.

7.3 Feature Based Birth Process

Although the $\omega_{\rm ML}$ configuration estimate does not depend on the birth maps, the exploration strategy in the population space affects the speed of optimization notably. In the bMBD algorithm (see Fig. 15) the most significant part of the computational time corresponds to calculating the $A_{\mathcal{D}}(u)$

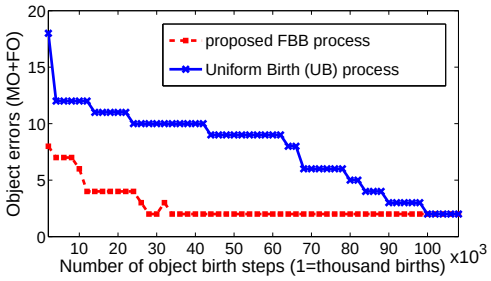


Fig. 23. Evolution of the detection performance over the iteration steps during optimization on the BUDAPEST image pair. Object level error is given as a function of the applied birth steps.

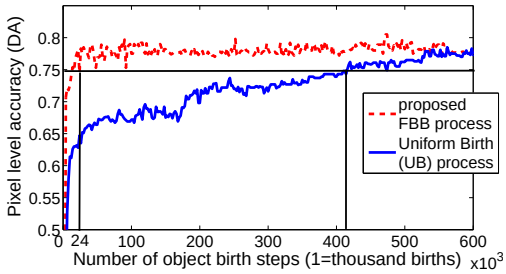


Fig. 24. Evolution of the detection performance over the iteration steps during the optimization on the BUDAPEST image pair. Detection Accuracy DA (i.e. pixel level F-score) is given as a function of the applied birth steps.

data term for each generated object, since the interaction potential $I(u, v)$ only needs to calculate the intersection area of rectangles which can be solved efficiently in an analytical way. For this reason, the complexity of the approach can be characterized by the number of object birth steps.

The key objective of the proposed non-uniform Feature Based Birth (FBB) procedure is to generate relevant objects with higher probability, so that we need to deal with less inefficient building segment candidates, and high quality configurations can be reached more quickly. We should note here that the $P_b^{(i)}(s)$, $P_{\mathrm{ch}}(s)$, $\mu_\theta^{(i)}(s)$, $\mu_L^{(i)}(s)$ and $\mu_l^{(i)}(s)$ *birth maps* are calculated only once before starting the iterative algorithm, and using dynamic programming techniques its computational need is negligible [28] considering the cost of the whole Simulated Annealing (SA) process.

For evaluation, we compared the convergence speed of the bMBD optimization using the proposed FBB and the conventional Uniform Birth (UB) processes. In the UB case, the $P_b^{(i)}(s)$ and $P_{\mathrm{ch}}(s)$ maps follow a uniform distribution and the side length/orientation parameters are also set as uniform random values. In Fig. 23, the object-errors are shown as a function of the birth steps: the FBB approach reaches the final error rate with 3 times less birth calls than the UB. The difference is even more significant at pixel level. As Fig. 24 shows, with the UB process the pixel level accuracy rates converge much slower than the object errors; to reach the 75% DA rate, we need to generate $400,000$ objects with the UB map, and only $24,000$ building candidates with the proposed FBB map. This observation means that the appearing object silhouettes in the uniform approach are usually notably inaccurate in the beginning, and considerable time is needed to reach the optimum.

8 CONCLUSION

We have proposed a multitemporal Marked Point Process (mMPP) framework for building extraction and change monitoring in remotely sensed image pairs taken with significant time differences. The method incorporates object recognition and low level change information in a joint probabilistic approach. A global optimization process attempts to find the optimal configuration of buildings, considering the observed

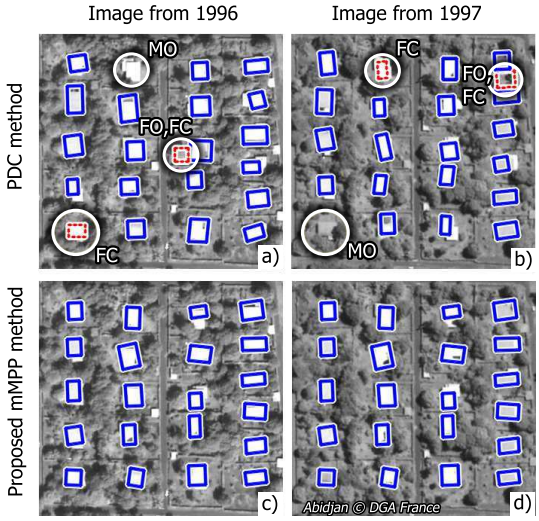


Fig. 25. Results on ABIDJAN images (source: DGA© France). Left: image from 1996, right: image from 1997. Top: Post Detection Comparison (PDC) (errors are highlighted by circles), Bottom: proposed joint mMPP model

data, prior knowledge, and interactions between the neighboring building parts. The accuracy is ensured by a Bayesian object model verification, meanwhile the computational cost is significantly decreased by a non-uniform stochastic object birth process, which proposes relevant objects with higher probability based on low-level image features.

9 ACKNOWLEDGEMENT

The authors acknowledge the test data provided by András Görög (BUDAPEST images), the French Defense Agency (ABIDJAN) and Véronique Prinet from LIAMA Laboratory of CAS Beijing (BEIJING). We are grateful to the authors of the reference methods and their colleagues for their help in the evaluation, especially to Dr. Beril Sirmaçek and Cosmin Mihai. We thank our colleagues at MTA SZTAKI for linguistic corrections and checking the notations of the manuscript.

Fig. 26. Results on BUDAPEST (top, image part - provider: András Görög) and BEIJING (bottom, provider: Liama Laboratory CAS, China) image pairs, marking the unchanged (solid rectangles) and changed (dashed) objects

TABLE 4
Quantitative evaluation results. #CH and #UCH denote the total number of changed resp. unchanged buildings in the set. PDC denotes the Post Detection Classification reference method and mMPP refers to the proposed multitemporal Marked Point Process model. Evaluation rates MO, FO, MC, FC and DA are introduced in Sec. 7.

			Missing Obj. (MO)		False Obj. (FO)		Missing Change (MC)		False Change (FC)		Pix. lev. F-score	
Data Set	#CH	#UCH	PDC	**mMPP**	PDC	**mMPP**	PDC	**mMPP**	PDC	**mMPP**	PDC	**mMPP**
BUDAPEST	20	21	3	0	7	2	1	0	9	2	0.72	0.78
BEIJING	13	4	1	0	2	1	0	0	3	0	0.77	0.85
SZADA	50	7	4	2	0	1	3	4	3	0	0.76	0.82
ABIDJAN	0	21	2	0	2	0	0	0	4	0	0.78	0.91

REFERENCES

[1] A. Katartzis and H. Sahli, "A stochastic framework for the identification of building rooftops using a single remote sensing image," *IEEE Trans. Geosc. Remote Sens.*, vol. 46, no. 1, pp. 259–271, 2008.

[2] F. Lafarge, X. Descombes, J. Zerubia, and M. Pierrot-Deseilligny, "Structural approach for building reconstruction from a single DSM," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 135–147, 2010.

[3] B. Sirmaçek and C. Ünsalan, "Building detection from aerial imagery using invariant color features and shadow information," in *Int. Symp. on Computer and Information Sciences (ISCIS)*, Istanbul, Turkey, 2008.

[4] S. Noronha and R. Nevatia, "Detection and modeling of buildings from multiple aerial images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 5, pp. 501–518, 2001.

[5] F. Bignone, O. Henricsson, P. Fua, and M. Stricker, "Automatic extraction of generic house roofs from high resolution aerial imagery," in *European Conf. on Computer Vision*, Cambridge, UK, 1996, pp. 83–96.

[6] M. Ortner, X. Descombes, and J. Zerubia, "A marked point process of rectangles and segments for automatic analysis of digital elevation models." *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 1, pp. 105–119, 2008.

[7] N. Champion, "2D building change detection from high resolution aerial images and correlation digital surface models," in *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2007, pp. 197–202.

[8] K. Karantzalos and N. Paragios, "Large-scale building reconstruction through information fusion and 3D priors," *IEEE Trans. Geosc. Remote Sens.*, vol. 48, no. 5, pp. 2283–2296, 2010.

[9] F. Rottensteiner, J. Trinder, S. Clode, and K. Kubik, "Building detection by fusion of airborne laser scanner data and multi-spectral images: Performance evaluation and sensitivity analysis," *ISPRS Journal for Photogrammetry and Remote Sensing*, vol. 62, no. 2, pp. 135–149, 2007.

[10] J. Jaw and C. Cheng, "Building roof reconstruction by fusing laser range data and aerial images," in *Proc. ISPRS Congress*, Beijing, China, 2008, pp. 707–712.

[11] J. Shufelt, "Performance evaluation and analysis of monocular building extraction from aerial imagery," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 4, pp. 311–326, 1999.

[12] F. Rottensteiner, "Automated updating of building data bases from digital surface models and multi-spectral images: Potential and limitations," in *ISPRS Congress*, Beijing, China, 2008, pp. 265–270.

[13] S. Tanathong, K. Rudahl, and S. Goldin, "Object oriented change detection of buildings after the Indian ocean tsunami disaster," in *IEEE International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, Krabi, Thailand, 2008, pp. 65–68.

[14] L. Bruzzone and D. Fernandez Prieto, "An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images," *IEEE Trans. on Image Processing*, vol. 11, no. 4, pp. 452–466, 2002.

[15] C. Benedek and T. Szirányi, "Change detection in optical aerial images by a multi-layer conditional mixed Markov model," *IEEE Trans. Geosc. Remote Sens.*, vol. 47, no. 10, pp. 3416–3430, 2009.

[16] A. Fournier, P. Weiss, L. Blanc-Fraud, and G. Aubert, "A contrast equalization procedure for change detection algorithms: applications to remotely sensed images of urban areas," in *International Conference on Pattern Recognition (ICPR)*, Tampa, FL, USA, 2008, CD–ROM.

[17] S. Müller and D. Zaum, "Robust building detection in aerial images," in *ISPRS Object Extraction for 3D City Models, Road Databases and Traffic Monitoring - Concepts, Algorithms and Evaluation, (CMRT05)*, Vienna, Austria, 2005, pp. 143–148.

[18] Z. Song, C. Pan, Q. Yang, F. Li, and W. Li, "Building roof detection from a single high-resolution satellite image in dense urban area," in *Proc. ISPRS Congress*, Beijing, China, 2008, pp. 271–277.
[19] X. Descombes and J. Zerubia, "Marked point processes in image analysis," *IEEE Signal Processing Magazine*, vol. 19, no. 5, pp. 77–84, 2002.
[20] C. Lacoste, X. Descombes, and J. Zerubia, "Point processes for unsupervised line network extraction in remote sensing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1568–1579, 2005.
[21] F. Lafarge, G. Gimel'farb, and X. Descombes, "Geometric feature extraction by a multi-marked point process," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1597–1609, 2010.
[22] X. Descombes, R. Minlos, and E. Zhizhina, "Object extraction using a stochastic birth-and-death dynamics in continuum," *Journal of Mathematical Imaging and Vision*, vol. 33, pp. 347–359, 2009.
[23] H. Hatsuda, "Automatic cell identification using a multiple marked point process," in *Int'l Conf. on Bioinformatics & Computational Biology*, 2010, pp. 605–609.
[24] Á. Utasi and C. Benedek, "A 3-D marked point process model for multi-view people detection," in *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs, USA, 2011, pp. 3385–3392.
[25] Z. Tu and S.-C. Zhu, "Image segmentation by Data-Driven Markov Chain Monte Carlo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, pp. 657–673, 2002.
[26] C. Benedek, X. Descombes, and J. Zerubia, "Building extraction and change detection in multitemporal remotely sensed images with multiple birth and death dynamics," in *IEEE Workshop on Applications of Computer Vision (WACV)*, Snowbird, USA, 2009, pp. 100–105.
[27] ——, "Building detection in a single remotely sensed image with a point process of rectangles," in *International Conference on Pattern Recognition (ICPR)*, Istanbul, Turkey, 2010, pp. 1417–1420.
[28] ——, "Building extraction and change detection in multitemporal aerial and satellite images in a joint stochastic approach," INRIA, Research Report 7143, December 2009, Available: http://hal.archives-ouvertes.fr/inria-00426615/en/.
[29] B. Sirmaçek and C. Ünsalan, "Urban-area and building detection using SIFT keypoints and graph theory," *IEEE Trans. Geosc. Remote Sens.*, vol. 47, no. 4, pp. 1156–1167, 2009.
[30] ——, "A probabilistic framework to detect buildings in aerial and satellite images," *IEEE Trans. Geosc. Remote Sens.*, vol. 49, pp. 211–221, 2011.
[31] P. Saeedi and H. Zwick, "Automatic building detection in aerial and satellite images." in *IEEE Intl. Conf. on Control, Automation, Robotics and Vision*, Hanoi, Vietnam, 2008, pp. 623–629.
[32] J. Porway, Q. Wang, and S. C. Zhu, "A hierarchical and contextual model for aerial image parsing," *Int. J. Comput. Vision*, vol. 88, no. 2, pp. 254–283, 2010.
[33] K. Khoshelham and Z. Li, "A split-and-merge technique for automated reconstruction of roof planes," *Photogrammetric Engineering and Remote Sensing*, vol. 71, no. 7, pp. 855–863, 2005.
[34] Z. Song, C. Pan, and Q. Yang, "A region-based approach to building detection in densely build-up high resolution satellite image," in *Intl. Conference on Image Processing*, Atlanta, Georgia, USA, 2006, pp. 3225–3228.
[35] K. Karantzalos and N. Paragios, "Recognition-driven two-dimensional competing priors toward automatic and accurate building detection," *IEEE Trans. Geosc. Remote Sens.*, vol. 47, no. 1, pp. 133–144, 2009.
[36] J. Peng, D. Zhang, and Y. Liu, "An improved snake model for building detection from urban aerial images," *Pattern Recognition Letters*, vol. 26, no. 5, pp. 587–595, 2005.
[37] S. Kumar and M. Hebert, "Detection in natural images using a causal multiscale random field," in *IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR)*, vol. 1, Madison, USA, 2003, pp. 119–126.
[38] V. Tsai, "A comparative study on shadow compensation of color aerial images in invariant color models," *IEEE Trans. Geosc. Remote Sens.*, vol. 44, no. 6, pp. 1661–1671, 2006.
[39] P. Dare, "Shadow analysis in high-resolution satellite imagery of urban areas," *Photogrammetric Engineering and Remote Sensing*, vol. 71, no. 2, pp. 169–178, 2005.

Csaba Benedek received the M.Sc. degree in computer sciences in 2004 from the Budapest University of Technology and Economics (BUTE), and the Ph.D. degree in image processing in 2008 from the Pázmány Péter Catholic University, Budapest. Starting from October 2008, he worked for 12 months as a postdoctoral researcher with the Ariana Project Team at INRIA Sophia-Antipolis, France. He is currently a senior research fellow with the Distributed Events Analysis Research Group, at the Computer and Automation Research Institute, Hungarian Academy of Sciences, and an assistant professor with the Dept. of Electronic Technology, BUTE. His research interests include Bayesian image segmentation and object extraction, change detection, video surveillance and remotely sensed data analysis.

Xavier Descombes received the bachelor's degree in telecommunications from the Ecole Nationale Superieure des Telecommunications (ENST) Paris, France, in 1989, the master of science in mathematics from the University of Paris VI in 1990, the PhD degree in signal and image processing from the ENST in 1993 and the "habilitation" in 2004 from the University of Nice Sophia-Antipolis. He has been awarded the "prix de la recherche" in human health in 2008. He has been a postdoctoral researcher at ENST in 1994, at the Katolieke Universitat Leuven in 1995, at the Institut National de Recherche en Informatique et en Automatique (INRIA) in 1996 and a visiting scientist in the Max Planck Institute of Leipzig in 1997. He is currently a permanent researcher at INRIA. His research interests focus on stochastic modeling in image processing.

Josiane Zerubia has been a permanent research scientist at INRIA since 1989, and director of research since July 1995. She was head of the PASTIS remote sensing laboratory (INRIA Sophia-Antipolis) from mid-1995 to 1997. Since January 1998, she has been head of the Ariana research group (INRIA/CNRS/University of Nice), which also works on remote sensing. She has been professor at SUPAERO (ISAE) in Toulouse since 1999.

She is a Fellow of the IEEE. She is a member of the IEEE IMDSP and IEEE BISP Technical Committees (SP Society). She was associate editor of IEEE Trans. on IP from 1998 to 2002; area editor of IEEE TRANS. ON IP from 2003 to 2006; guest co-editor of a special issue of IEEE TRANS. ON PAMI in 2003; and member-at-large of the Board of Governors of the IEEE SP Society from 2002 to 2004. She has also been a member of the editorial board of the French Society for Photogrammetry and Remote Sensing (SFPT) since 1998, of the International Journal of Computer Vision since 2004, and of the Foundation and Trends in Signal Processing since 2007. She has been associate editor of the on-line resource: Earthzine (IEEE CEO and GEOSS). She was co-chair of two workshops on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR'01, Sophia Antipolis, France, and EMMCVPR'03, Lisbon, Portugal); co-chair of the special sessions at IEEE ICASSP 2006 (Toulouse, France) and at IEEE ISBI 2008 (Paris, France). She is a member of the organizing committees of IEEE ICIP 2011 (Brussels, Belgium) and IEEE ICIP 2014 (Paris, France).

Her current research interests are in image processing using probabilistic models and variational methods. She also works on parameter estimation and optimization techniques.